INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

NOTE TO USERS

This reproduction is the best copy available.

UMI

DELAY ESTIMATION AND ITS QoS IMPLICATIONS IN VOICE OVER IP NETWORKS

Submitted by

MARIOS S. PARPERIS

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master's of Electrical Engineering

**Ottawa-Carleton Institute for Electrical Engineering
Faculty of Engineering
Department of Systems and Computer Engineering**

**Carleton University
Ottawa, Ontario, Canada, K1S 5B6
January 25th 2000**

**© Copyright
2000 Marios S. Parperis**

**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-52394-2

Canada

The undersigned hereby recommended to
the Faculty of Graduate Studies and Research
acceptance of the thesis

**DELAY ESTIMATION AND ITS QoS IMPLICATIONS IN
VOICE OVER IP NETWORKS**

Submitted by
Marios S. Parperis

in partial fulfillment of the requirements
for the degree of Master's of Electrical Engineering

Dr. Ioannis Lambadaris, Thesis Supervisor

Dr. Rafik Goubran, Thesis Co-Supervisor

Dr. Rafik Goubran, Chair, Department of Systems and Computer Engineering

Carleton University

February 1st, 2000

ABSTRACT

Over the past few years Computer telephony integration (CTI) has gained significant popularity on the Internet and Intranet community. CTI integrates voice communication technology such as voice over Internet protocol (VoIP) with the ability to exchange data, whiteboards and videoconferencing. This thesis presents original new structures that enhance the capabilities of VoIP platforms.

Two research oriented VoIP platforms were developed in order to demonstrate acoustic round trip delays measurements and determine acoustic echo suppression targets. The platforms are H.323 compliant. The real time protocol and real time control protocol (RTP/RTCP) were incorporated in the two platforms.

A theoretical model and an empirical measurement method for the acoustic round trip delay (ARTD) of a VoIP conference is proposed in this thesis. An acoustic domain verification test is also used to check the ARTD empirical measurements for accuracy.

This thesis also proposes a method for determining acoustic echo cancellation targets by combining acoustic round trip delay results and psycho-acoustic echo target data. Accurate determination of echo suppression targets is required to minimize detrimental switched loss attenuation effects.

ACKNOWLEDGEMENTS

I would like to express my thanks to my supervisors and colleagues for their assistance and support during the development of my thesis. Dr. Ioannis Lambadaris and Dr. Rafik Goubran for their guidance and support in this research project; Trevor Yensen for his role in developing the Multi-Participant Intranet Phone System (MIPS) and his participation in the development of the acoustic round trip delay measurement method both the empirical and the theoretical model.

I would also like to thank my wife Kate Parperis for her many years of encouragement and for her hard work in editing this thesis.

I would like to thank my family for their many years of moral and financial support and encouragement during my university carrier.

I would also like to thank Newbridge Networks Corporation and the Ontario Graduate Scholarship funds for their financial support.

Last but not least, the support staff at Carleton university for their assistance: Danny Lemay and the other lab technicians for their help; the secretarial staff in the Department of Systems and Computer Engineering for all their help.

TABLE OF CONTENTS

1	INTRODUCTION.....	1
1.1	THESIS OBJECTIVES.....	1
1.2	CONTRIBUTIONS TO ENGINEERING KNOWLEDGE.....	4
2	BACKGROUND INFORMATION.....	6
2.1	REAL TIME TRANSPORT PROTOCOL (RTP).....	6
2.2	MAIN TASKS OF THE REAL TIME PROTOCOL	7
2.3	REAL TIME PROTOCOL SPECIFIC FUNCTIONS	9
2.4	RESYNCHRONIZATION	10
2.5	INTER-ARRIVAL JITTER.....	13
2.6	RTCP PACKET FORMAT	15
2.7	RTP HEADER EXTENSION	17
2.8	INTERNET PROTOCOL	18
3	DESIGNING OF A GENERIC PLATFORM.....	22
3.1	HARDWARE REQUIREMENTS	23
3.2	SOFTWARE REQUIREMENTS.....	23
3.3	AUDIO DATA FORMAT.....	24
3.4	H.323 SOFTWARE SUPPORT	24
3.4.1	<i>Connectionless Service Requirement.....</i>	<i>26</i>
3.5	MULTICASTING SUPPORT	28
3.6	SOFTWARE DESIGN.....	30

4	PROPOSED METHOD OF CALCULATING ROUND TRIP DELAY	33
4.1	INTRODUCTION	34
4.1.1	<i>Testing Platform</i>	35
4.1.2	<i>Quality of Service</i>	35
4.1.3	<i>Echo Suppression</i>	36
4.1.4	<i>Acoustic Echo Cancellation</i>	36
4.2	PSYCHOACOUSTIC DETERMINATION OF ECHO TARGET.....	39
4.3	THEORETICAL MODEL OF ACOUSTIC ROUND TRIP DELAY	41
4.3.1	<i>Explanation of Delays</i>	41
4.3.2	<i>Pipeline Analogy</i>	46
4.4	THEORETICAL ARTD EQUATION	48
5	RESULTS	50
5.1	INTRODUCTION	50
5.2	EMPIRICAL MEASUREMENT METHOD FOR ACOUSTIC ROUND TRIP DELAY	50
5.2.1	<i>Empirical Measurement method Results</i>	55
5.3	ACOUSTIC DOMAIN VERIFICATION OF ACOUSTIC ROUND TRIP DELAY	58
5.3.1	<i>Calculations</i>	61
5.4	ECHO SUPPRESSION TARGET DETERMINATION	64
5.5	CONCLUSIONS	67
6	CONCLUSIONS – FUTURE WORK.....	69
6.1	FUTURE RESEARCH	70

7 REFERENCES.....	71
APPENDIX A Software Design	76
APPENDIX B Hardware/Software requirements for dMIPS	90

LIST OF FIGURES

FIGURE 1 – RTP HEADER FORMAT	8
FIGURE 2 – FRAME VALIDATION FLOWCHART	11
FIGURE 3 – RTCP PACKET FORMAT (SENDER REPORT).....	17
FIGURE 4 - IP PACKET FORMAT.....	19
FIGURE 5 - HIGH LEVEL VIEW OF VOIP PLATFORM.....	22
FIGURE 6 - HIGH-LEVEL VIEW OF H.323 TOOLKIT STANDARD.....	25
FIGURE 7 - TCP/IP PROTOCOL STACK.....	27
FIGURE 8 - SCREENSHOT OF DMIPS	31
FIGURE 9 - HIGH-LEVEL VIEW OF H.323 TERMINAL EQUIPMENT	34
FIGURE 10 – ACOUSTIC ECHO CANCELLATION STRUCTURE.....	37
FIGURE 11 – ACOUSTIC ECHO CANCELLATION STRUCTURE WITH SWITCHED-LOSS.....	38
FIGURE 12 – TALKER ECHO TOLERANCE CURVES.....	40
FIGURE 13 – HIGH LEVEL DIAGRAM FOR THE ACOUSTIC ROUND TRIP DELAY	42
FIGURE 14 – PIPELINE ANALOGY MODEL OF THE ARTD PATH	47
FIGURE 15 – EMPIRICAL MEASUREMENT OF ACOUSTIC ROUND TRIP DELAY	51
FIGURE 16 - DSP ROUND TRIP DELAY MEASUREMENTS FOR 20MS FRAME	56
FIGURE 17 - DSP ROUND TRIP DELAY MEASUREMENTS FOR 30MS FRAME.	57
FIGURE 18 - DSP ROUND TRIP DELAY MEASUREMENTS FOR 40MS FRAME	58
FIGURE 19 - VERIFICATION MODEL.....	59
FIGURE 20 – CROSS CORRELATION OF WGN FOR $T_F=20\text{MS}$	60
FIGURE 21 – CROSS CORRELATION OF WGN VECTORS FOR $T_F=40\text{MS}$	61
FIGURE 22 – VERIFICATION RESULTS FOR $T_F=20\text{MS}$	63

FIGURE 23 - VERIFICATION RESULTS FOR $T_f=40\text{MS}$	64
FIGURE 24 - PSYCHO-ACOUSTIC DETERMINATION OF SWITCHED-LOSS	65
FIGURE 25 - COMBINED ARTD AND ECHO TOLERANCE CURVES	67
FIGURE 26 - USE CASE MAP SYMBOLS	76
FIGURE 27 - HIGH-LEVEL USE CASE MAP FOR RTP/DSP SOFTWARE	77
FIGURE 28 - JITTER BUFFER USE CASE MAP	78
FIGURE 29 - SEND DATA USE CASE MAPS	79
FIGURE 30 - RECEIVE DATA USE CASE MAP	81
FIGURE 31 - JITTER BUFFER STRUCTURE	83
FIGURE 32 - HOST PORT INTERFACE AND TRANSFER REGISTERS	84
FIGURE 33 - CAPTURE AND JITTER BUFFER	85
FIGURE 34 - ROUND TRIP DELAY INFORMATION	86
FIGURE 35 - TYPICAL CONTROL USE CASE MAP	87

LIST OF ACRONYMS

ABS	Audio Buffer Delay
ADPCM	Adaptive Differential Pulse Code Modulation
AEC	Acoustic Echo Canceller
ARTD	Acoustic Round Trip Delay
CTI	Computer Telephony Integration
DMA	Direct Memory Access
dMIPS	DSP Multiparticicipant Intraphone System
DSP	Digital Signal Processor
EABD	Expected Audio Buffer Delay
EGP	Exterior Gateway Protocol
ERLE	Echo Return Loss Enhancement
IPng	Internet Protocol Next Generation
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
ITU	International Telecommunication Union
JBS	Jitter Buffer Delay
LAN	Local Area Network
MBone	Multimedia Backbone
NLP	Non Linear Processor
PC	Personal Computer
PCM	Pulse Code Modulation
PDU	Protocol Data Unit

PIO	Programmed Input Output
POTS	Plain Old Telephone Service
PSTN	Public Service Telephone Network
QoS	Quality of Service
RTCP	Real Time Control Protocol
RTP	Real Time Protocol
SSRC	Synchronization Source
TCP	Transport Control Protocol
TDPT	Total Delay Playout Threshold
TELAR	Talker Echo Loudness Rating
TTL	Time to Live
UDP	User Datagram Protocol
VoIP	Voice over Internet Protocol

CHAPTER 1

1 INTRODUCTION

The information superhighway is gearing to serve as a vehicle for the exchange of rich multimedia data such as audio and video along with traditional data types such as text and still images [32]. Voice over Internet protocol (VoIP) is an important new technology that is being considered by many companies as future competition for public service telephone networks (PSTN). VoIP applications will be combined with many new multimedia services not currently available over PSTNs such as videoconferencing, white boards, fax and other data exchange capabilities.

VoIP systems must provide better hands free full-duplex conferencing support for users to compete with PSTN service. Speakerphones are in wide use by business customers for conference calling and by home users for the convenience offered from having their hands free to perform other activities. Furthermore, full-duplex capabilities are essential since it has been determine that standard conversation protocols are disrupted by half duplex systems. The delay in half duplex switching can cause a talker's speech to be cut off and create switching noise.

1.1 Thesis Objectives

A practical implementation of a full-duplex conferencing system involves the design of a control system required to regulate half-duplex attenuation effects. This will provide the high level of echo cancellation required. Half-duplex effects are called switched-loss attenuation and a nonlinear processor (NLP) could compute the amount of attenuation to be

applied in a conference. It is evident that these half-duplex effects should be minimized in order for the user to perceive the conference as full duplex. Psycho-acoustic data has shown that the level of total echo cancellation required may be determined by the mean one-way propagation delay of the system. Conventional public service telephone networks (PSTN) are unable to determine this parameter, however VoIP systems are capable of measuring this value.

In a practical VoIP system implementation, high levels of acoustic echo cancellation are required. Switched loss, which is applied to the signal path in addition of echo cancellation, is needed. A control system (NLP) is required to generate switch loss. The level of switched loss should be minimized in order to maintain full-duplex conferencing without the negative half-duplex effects. A new algorithm to minimize half-duplex effects is proposed in this thesis by determining echo return loss enhancement (ERLE) through psycho-acoustic data. ERLE is based on the mean one-way propagation delay for a voice conference. The mean one way propagation delay may be roughly approximated to one half of the round trip delay.

A method for determining the mean one-way propagation delay of the system must be determined in order to compute the amount of echo cancellation required. Mean one-way propagation delay may be approximated through the determination of the complete acoustic round trip delay as perceived by the user and then dividing it by two. In order to determine the complete round trip delay of a VoIP system, a complete investigation into the delays associated with the proposed conferencing systems was undertaken. An empirical

measurement system was implemented on the Carleton University DSP/PC VoIP platform and a multimedia PC platform.

A Carleton University VoIP test platform was developed in order to retrieve acoustic round trip delay data. There are no commercially available systems that can be modified to support our research, so a platform was created at Carleton University. The platform is an n-way VoIP system using a PC with a DSP card. It was named DSP Multi-participant Intranet Phone System (dMIPS). It is capable of multiple participant conferencing (n-way) and conforms to the H.323 specification for compatibility with the emerging VoIP standard. The DSP platform is of interest since it has the processing power to support acoustic echo cancellation and is capable of much lower round trip delays than other platforms based on PC sound card technology.

This thesis proposes that there is an acceptable solution to a multiple participant stereo conferencing platform. A thorough investigation into the delays associated with VoIP conferencing was performed and allowed us to develop an empirical method of measuring the acoustic round trip delay of the system. With knowledge of the round trip delay of the system we are able to determine suitable acoustic echo cancellation targets from psycho-acoustic data. Furthermore, the design of the dMIPS platform is discussed in detail and its conformance with the dominant standard H.323 is emphasized. The information provided will present a complete solution to the problem of designing an appropriate acoustic interface to a VoIP system.

1.2 Contributions to engineering knowledge

This thesis proposes a theoretical model and an empirical method for measuring the acoustic round trip delay in a VoIP conference. The acoustic round trip delay is the sum of all delays from the time the near end user begins talking until the far end echo - which is generated at the far end - returns to user's ear. An empirical measurement method is proposed to validate proposed equations. The empirical measurement method was implemented on the Carleton University's digital signal processor/personal computer (DSP/PC) platform and multimedia PC platform to collect acoustic round trip delay data.

A Carleton University VoIP platform was developed to measure acoustic round trip delay through the acoustic round trip delay empirical measurement method. A custom platform was developed since there were no commercially available systems capable of supporting the proposals of this thesis. The platform conforms to H.323 standard for compatibility with other VoIP conference systems. Two sub-platforms were developed: A DSP/PC based which provides higher processing power and a multimedia PC platform to test acoustic round trip delay on sound card based systems.

In summary, the contributions of this thesis on the engineering knowledge are:

1. Design and implement an empirical method for measuring acoustic round trip delays in VoIP systems.
2. Relate acoustic round trip delay data to psycho-acoustic research data showing echo suppression targets.

3. A multimedia PC platform (multi-participant Intranet phone system, MIPS) was designed and implemented at Carleton Univesrity DSP laboratory to test a proposed acoustic round trip delay measurement method in CTI systems.
4. A multimedia PC/DSP platform (dMIPS) was also designed and implemented. This platform provided a high level of processing power necessary for acoustic echo cancellation in CTI systems.

CHAPTER 2

2 BACKGROUND INFORMATION

This chapter provides some essential background information on protocols and techniques used to develop the Carleton University Voice over IP platform.

2.1 Real time Transport protocol (RTP)

RTP is a transport protocol that provides end to end services for data with Real Time characteristics such as interactive audio and video [19]. Those services require payload and type identification, sequence numbering, time stamping and delivery monitor. RTP is running on the top of UDP/IP. Both protocols contribute parts of the transport protocol functionality. RTP also supports data transfer to multiple destinations using multicast distribution if provided by the underlying network.

RTP represents a new style protocol following the principles of application level framing and integrated layer processing proposed by [5]. That is, RTP is intended to be able to provide the information required by a particular application and will often be integrated into the application processing rather than being implemented as a separate layer. In conventional protocols making the protocol more general might accommodate additional functions. On the contrary, RTP is intended to be tailored through modifications and/or additions to the headers of each frame as needed.

2.2 Main Tasks of the Real Time Protocol

The main tasks of RTP is to:

- Specify and check a payload type
- Create/check/update the sequence number
- Validate a frame
- Generate/update statistics

Statistics may include cumulative frames/octets lost, interarrival jitter, sequence monitoring as well as the timestamp of each individual packet. Using the above statistics you may extract even more useful information such as the acoustic round trip delay, and the number of lost packets over a certain period of time.

Each participant sends audio data in small packets. Each packet of audio data is preceded by an RTP header. RTP header and data are in turn contained in a UDP packet. The RTP header indicates what type of audio encoding (such as PCM, ADPCM) is contained in each packet so that senders can change the encoding during a conference. For example, each sender may accommodate a new participant that is connected through a low bandwidth link or react to indications of network congestion.

The Internet, like other packet networks, occasionally loses and re-orders packets and delays them by variable amounts of time. To cope with these impairments, the RTP header contains timing information and a sequence number that allow the receivers to reconstruct the timing produced by the source, so that packets are continuously played out of the speaker. This timing reconstruction is performed separately for each of the RTP packets in the conference.

The sequence number can also be used by the receiver to estimate how many packets are lost. The RTP header has the format that is shown in Figure 1.

V	P	Ext.	Contr. Src	M	P T	Seq. Number
Timestamp						
Synchronization Source Identifier						
Contributing Source Identifiers ...						

Figure 1 – RTP Header Format

The various fields of the RTP header are as follows:

V: Version (2bits): This is the field that identifies the version of RTP. The version defined in this thesis is 2.

P: Padding (1 bit): If the padding is set, the packet contains one or more additional padding octets at the end which are not part of the payload. The last octet of the padding contains a count of how many padding octets should be ignored.

Ext: Extension (1 bit): If the extension bit is set, the fixed header is followed by exactly one header extension.

Contr. Src: Contributing Source Count. (4 bits): Identifies the number of participants in a conference.

M: Marker bit (1 bit): It is intended to allow for significant events like frame boundaries to be marked on a frame stream.

P T: Payload Type (7 bits): This field identifies the format of the RTP payload and determines its interpretation by the application.

Seq. Number: Sequence Number (16 bits): Increments by one for each RTP data packets sent and may be used by the receiver to detect packet loss.

Timestamp (32 bits): The timestamp reflects the sampling instant of the RTP data packet.

Synchronization Source Identifier (32 bits): This field identifies the source. In our case, the IP address of the source was used.

Contributing Source Identifiers (x32 bits): IP addresses of the contributing sources.

2.3 Real Time protocol Specific Functions

In order to efficiently present the various RTP functions a separation of the two ends, Sender and Receiver, is needed.

On the sender part the protocol is responsible for creating a header for each frame and encapsulate the data received from the application/session layer. Among the parameters that have to be set are the version of RTP, the padding flag, specification whether there is a header extension or not and setting of the marker bit (signifies important events on the application layer, such as frame boundaries in packet streams). Also, the payload type, the sequence number, the time stamp, and finally the synchronization source (IP Address) have to be set.

RTP is designed to allow an application to scale automatically over session sizes ranging from a few participants to, potentially, hundreds. For example, in an audio conference, the data traffic is inherently self-limiting because one or two people could speak at a time. Furthermore, with multicast distribution the data rate at any given link remains relatively

constant, independent of the number of participants. However, the control traffic does not remain constant. It is rather dependent on the number of users currently active in a conference call. If the reception reports from each were sent at a constant rate, the control traffic would grow linearly with the number of participants. Therefore the rate of the control packets should be scaled according to the number of participants.

On the other end, the receiver has to decapsulate the data packet and interpret the header information. The receiver has to check for payload type, version and validate the frame. After that it has to calculate the statistics (jitter buffer, delay packets lost etc) and update them. Furthermore, the receiver must be able to resynchronize itself if the participant was active for a certain period of time. Resynchronization is needed in case there is no session control protocol present in the OSI stack.

2.4 Resynchronization

Not all Real Time Protocols can afford frame correction or reordering. For example, if there is network delay greater than the buffer capacity of a system, any delayed frames must be discarded. Otherwise, the session will lose from its real time nature. Furthermore, the possible lack of a session protocol leaves open the possibility that inactive participants may exist.

Resynchronization of a participant in a conference, is the process of zeroing all statistics and starting a new session with this participant. This may occur when a participant reports significant amount of lost frames or if she/he stops transmitting frames for a certain amount

of time. The amount of time or frames lost before a participant need to be resynchronized (resynchronization threshold) may be a user-configured parameter.

A practical example is when a participant is not sending any data for some period of time and then rejoins the session. Then, resynchronization is necessary. Resynchronization is also useful in order to update correctly the statistics.

When re-synchronization occurs, the last frame received is treated as the first frame received from that synchronization source. All the statistics are zeroed and the last frame received is given a valid status.

Figure 2 is depicting a flowchart explaining the logic for all the possible scenarios when dealing with (re)synchronization.

Figure 2 – Frame Validation Flowchart

Referring to Figure 2, when a frame is received, its sequence number is read and a frame validation procedure is followed. The scope of the above flowchart is to determine if this frame is outdated (in that case discard or resynchronization is necessary) or if it is a valid one and it should be further processed.

Block 1 checks the sequence number of the current frame and compares it to the sequence number of the previous frame. If current sequence number is greater, then block 2 checks if the distance between the two is within a pre-specified threshold. If yes, then frame is OK, if not then there are only two possible scenarios: Either the frame is outdated but valid or frame is older than the current one (wrap around case). In the first case, resynchronization is initiated and in the newly arrived, frame is discarded.

Back to block 1, if the current sequence number is less or equal to the last one, then we have 4 cases:

1. Current sequence number is equal to the previous frame sequence number, in that case frame is discarded,
2. The case that we have a wrap around where the current frame is within the threshold, frame is OK.
3. Wrap around where the current frame is outside the threshold (i.e. participant being idle more than the threshold), resynchronization is needed.
4. And finally, the current sequence number is older than the previous one; this signifies a delayed frame that must be discarded.

2.5 Inter-arrival Jitter

Jitter is defined as the deviation in the inter-frame arrival times induced by random variations in media access and queuing delays at intermediate nodes [10]. Resource contention at the sending and receiving ends may also contribute to jitter.

Let S_i is the RTP timestamp from packet i and R_i , is the time of arrival (RTP timestamp) for packet i , then for two packets i and j . Then, the mean deviation of the difference in time spacing at the receiver compared to the sender for a pair of packets, $D_{(i,j)}$, may be expressed as:

$$D_{(i,j)} = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \quad (2-1)$$

Interarrival jitter is an estimate of the statistical variance of the RTP data packet interarrival time, measured in timestamp units. These units are user/system defined. The most common units used are milliseconds. RTP and RTCP layers provide this timestamp information that may be used to extract useful statistical information about an ongoing session such as round trip delay etc. The interarrival jitter, J , is defined to be the mean deviation (smoothed absolute value) of the difference, D , in packet spacing at the receiver compared to the sender for a pair of packets. As shown in equation (2-1), this is also equivalent to the difference in the “relative transit time” for the two packets; the relative transit time is the difference between a packet’s RTP timestamp and the receiver’s clock at the time of arrival, measured in the same units.

The interarrival jitter may be calculated as each data packet i , is received from a source, using the absolute difference for that packet and the previous packet, $i-1$, $D_{(i, i-1)}$ according to equation (2-2):

$$J_i = J_{i-1} + (|D_{i-1,i}| - J_{i-1}) \quad (2-2)$$

Where J_i and J_{i-1} are the interarrival jitters for packets i and $i-1$, respectively.

Whenever a reception report is issued, the current value of J_i is sampled.

When the above calculation is carried out for packets/frames that go through the Internet, a precision adjustment is needed. Consequently, every time a new Jitter is calculated may also be divided by 16. Hence, equation (2-2) becomes:

$$J_i = \frac{J_{i-1} + (|D_{(i-1,i)}| - J_{i-1})}{16} \quad (2-3)$$

The jitter calculation is prescribed here to allow profile independent monitors to make valid interpretations of reports coming from different implementations. The above algorithm is the optimal first-order estimation and the gain parameter $1/16$ gives a good noise reduction ratio while maintaining a reasonable rate of convergence of the Jitter [15].

2.6 RTCP Packet Format

The real time control protocol (RTCP) is based on periodic transmission of control packets to all participants in a conference using the same distribution mechanism as the data packet [19]. The underlying protocol provides multiplexing of the data and control packets using different port numbers in UDP. RTCP performs four functions.

- The primary function is to provide feedback on the quality of data distribution. This is an integral part of the real time protocol's role as a transport protocol and is related to the flow and congestion control functions of other transport protocols. This feedback may be directly useful for control or adaptive encoding [16], [14]. Experiments with IP multicasting have shown that it is also critical to get feedback from the receivers to diagnose faults in the distribution. Reception feedback reports are originating from all participants. Real time control protocol helps a user who observes problems (delays or lost frames) to evaluate whether those problems are local or are involving more than one user. With a distribution mechanism like IP multicast, it is also possible for an entity such as a network service provider who is not otherwise involved in the session to receive the feedback information and act as a third-party monitor to diagnose network problems.
- RTCP carries a persistent transport-level identifier for each RTP source called the canonical name (Canonical end point identifier, CNAME). The Synchronization Source (SSRC) identifier of a user (which may be the IP address of a participant) may change if a conflict is discovered or a program is restarted. Therefore, receivers require the distinct

name (CNAME) to keep track of each participant. Receivers also require this distinct name to associate multiple data streams from the same participant in a set of related RTP sessions, for example to synchronize audio and video.

- The first two functions require that all participants send RTCP packets. Consequently, the rate must be controlled in order for the real time protocol to be able to accommodate a large number of participants. Each user sends its control packets to all the others. Hence, each can independently observe the number of participants in a conference call. The number of participants in a conference call may be used to calculate the rate at which the packets are sent.
- A fourth, optional function is to convey minimal session control information. This may include just participant identification. This is most likely to be useful in "loosely controlled" sessions where participants enter and leave without membership control or parameter negotiation. RTCP serves as a convenient channel to reach all the participants, but it is not necessarily expected to support all the control communication requirements of an application. A higher-level session control protocol, which is beyond the scope of this thesis, may be needed.

RTCP packet (Sender Report) format can be seen in Figure 3.

V	P	RC	P T = SR = 200	Length
SSRC of Sender				
NTP timestamp (MSW)				
NTP timestamp (LSW)				
RTP Timestamp				
Sender's packet count				
Sender's octet count				
Fraction lost		Cumulative packet lost		
Extended Highest Sequence Number Received				
Interarrival Jitter				
Last Sender report				
Delay since Last Sender report				

Figure 3 – RTCP Packet Format (Sender Report)

Where, V is the version of the protocol, P is the padding bit, RC is the receiver report count, PT is the packet type and NTP timestamp is the absolute time and is represented using the timestamp format of the Network Time Protocol (NTP) relative to 0h UTC on 1 January 1900 [8].

2.7 RTP Header Extension

An extension mechanism is provided to allow individual implementations to experiment with new payload format-independent functions. These functions may require additional information to be carried in the RTP data header [19]. This mechanism is designed so that the header extension may be ignored by other inter-operating implementations that have not being extended. In order to add an extension to an RTP packet, the extension bit on the RTP

header must be set (see Figure 1). When this bit is set, the fixed header is followed by exactly one header extension. The RTP header extension was used to develop the empirical model that in turn was used to calculate the acoustic round trip delay.

2.8 Internet Protocol

The Internet protocol (IP) is responsible for packet fragmentation and reassembly, routing and error reporting. IP provides a best-try connectionless service. The version of IP most widely used is version 4 (IPv4). IP version 6 (IPv6) is the next generation Internet protocol. It is also called IP next generation (IPng) (see 2.8.1). Ipv4 packet format is shown in Figure 4.

Packet information and reassembly are resource-consuming operations that occur when an IP packet is too large to be transmitted over the desired link and must be split into several smaller packets. Avoiding fragmentation by intermediate systems is desirable for real time

traffic to avoid extra overhead. Real time streams should set the "don't

Figure 4 - IP Packet Format

fragment" bit in the IP header to ensure that the network chosen to route the IP packet can handle datagrams of that size.

A specific type of service can be requested in the IP header to influence route selection. A priority 0 to 7 and flags to indicate low delay, high throughput or high reliability can set to influence the type of service [13].

The time-to-live (TTL) for an IP datagram is specified in its header and is decremented at each router along its route. This quantity can be used to limit the time the datagram is in transit across the network, since it will be discarded when the TTL reaches zero. Severely delayed datagrams cannot usually be covered and should be discarded.

Gateways are used to route IP datagrams if they are not addressed to the local area networks (LAN) [20]. There are two types of gateways that use different routing protocols. An interior gateway runs the interior gateway protocol (IGP). IGP maintains sufficient information to forward datagrams to hosts or connected LANs or to other interior gateways within the same autonomous system [13]. An autonomous system refers to a network that is separately managed and run. An exterior gateway runs the exterior gateway protocol (EGP) [34][35]. EGP maintains sufficient information to forward a datagram to an interior gateway if the datagram is destined for the same autonomous system or another exterior gateway if the datagram is destined for another autonomous system.

2.8.1 Internet Protocol Version 6 (IPv6)

IPv6 expands routing and addressing capabilities. The address fields have increased from 32 bits to 128 bits to support the rapid growth of the Internet and its addressing hierarchy.

Greater support for multicasting has been introduced by replacing the broadcast address class with a multicast address class.

A new address type called “any-cast” has been defined to specify that a datagram is destined for one of a set of nodes. Ipv6 defines specific values for its 4-bit priority field. A priority value between 0-7 indicates traffic for which the source will provide congestion control and 8-15 indicates traffic that does not back off in response to congestion such as real time traffic at a constant bit rate. The IPv6 header format has been simplified by eliminating fields, which are less used [35][37]. These simplifications were necessary to offset the increased header size due to the expanded address fields.

IPv6 supports new fields to specify the desired quality-of-service (QoS). Datagrams are grouped into traffic flows and the sender can request special handling for a real-time service. IPv6 contains many enhancements that should be used to enhance future VoIP products.

CHAPTER 3

3 Designing of a Generic Platform

The Carleton University real-time VoIP platform was designed to test various network issues in Voice over IP. These issues include the calculation of the acoustic round trip delay in a generic Voice over IP system that may lead to the estimation echo cancellation targets. The generic platform also helps to extract useful statistics that may help to improve QoS (Quality of Service) of the overall system.

Figure 5 - High Level View of VoIP platform

These statistics may be computed through the real time control RTCP and may involve lost frames, network jitter, timestamps of each frame and acoustic round trip delay for each of the participants. We named the VoIP platform “DSP multi-participant Intranet phone system” (dMIPS). The high level diagram of this system can be seen in Figure 5

3.1 Hardware Requirements

The PC platform that is required to support the dMIPS software includes a Pentium Class PC system or greater, Windows NT Workstation 4.0 or Windows '95/98, visual C++ V5.0 32-bit, PCI 10Mb/s Ethernet card, Winsock 2.0, Driver X software, multimedia speakers and microphone, microphone pre-amplifier and DSP board and software. Analytically, all the components needed may be found in Appendix B.

3.2 Software Requirements

The software resident on the DSP card performs the acoustic echo cancellation, spatialization, audio coding and decoding and jitter buffering required by the platform. The PC provides the Ethernet network interconnection by sending and receiving packets, performing RTP encapsulation and transferring data frames to and from the DSP card. The software that run on the DSP and PC platforms communicates through shared memory on the DSP card, called the host port interface. Interaction between the DSP and the PC involves the transfer of registers and interrupts requests.

The DSP board generates interrupts to the PC whenever captured data is ready for transmission by the PC. The PC generates interrupts to the DSP whenever valid data frames

are received by the network connection. The data is transferred by programmed I/O (PIO) mode, instead of direct memory access (DMA) mode, because the Tiger 542/PC does not support DMA. Using a DSP board capable of DMA transfers would reduce this data transfer overhead.

3.3 Audio Data Format

The software supports a simple audio format. The codec passes linear PCM 16-bit data at a fixed bit-rate of 8kHz. Plain old telephone service (POTS) telephone standard is using 8kHz 8-bit data streams; however to enhance the echo cancellation depth and sound quality the platform employs a 16-bit word length.

3.4 H.323 Software Support

MIPS and dMIPS platforms conform to the dominant (for CTI systems) standard H.323. The standard does not require that specific audio or video equipment should be used. This is left to the discretion of the designer but full compliance with H.323 ensures that this equipment will be inter-operable. The network interface is neither specified, nor are there any requirements for the support of mandatory data applications, such as white boards. H.323 is called a "toolkit" standard since it combines several ITU-T standards. This is illustrated in Figure 6. The shaded squares of the figure show the blocks that were implemented for the Carleton University dMIPS VoIP platform.

H.323 Recommendation

Figure 6 - High-Level View of H.323 Toolkit standard

The H.323 standard is independent of network topology for providing a maximum level of customization, and for easy implementation on existing networks. Voice is a required service to be carried by all H.323 compliant networks. Video and data may be supported optionally. The following levels of multimedia services are available through H.323:

Voice only	(mandatory support is required)
Voice and video	(video is optional)
Voice and data	(data is optional)
Voice, video and data	(video and data are optional)

The standard is intended to be as open as possible to support innovation by manufacturers and researchers. Terminals from the same manufacturer may support custom feature sets. If one of the terminals involved in the conference does not provide the proprietary feature set,

then a fallback onto mandatory components specified in H.323 will ensure conferencing is still possible.

The multiple participant conferencing platform is capable of operating in real-time on a PC. Data transmission over the network conforms to the recommendations laid out in the ITU-T H.323 [25] concerning the usage of RTP/RTCP protocols.

The functions provided by our implementation of the RTP/RTCP protocol, ITU-T H.225.0, [24], are as follows:

- end-to-end delivery of data with real-time characteristics
- payload type identification for each frame
- sequence numbering
- timestamping for each audio frame
- delivery monitoring
- quality of service monitoring through RTCP. This may include the cumulative number of lost frames, network jitter, timestamp information for each packet and acoustic round trip delay computation.

3.4.1 Connectionless Service Requirement

The RTP/RTCP data frames may be carried using the TCP/IP or UDP/IP protocols. The transmission control protocol/Internet protocol (TCP/IP) suite is the most widely used of protocols for guaranteed quality of service (QoS) networks. The TCP/IP suite provides reliable packet transmission service through the transmission control protocol (TCP) and

unreliable best-try service through the user datagram protocol (UDP). The function of the IP layer is to route TCP packets or UDP datagrams. Figure 7 shows a hierarchical view of the TCP/IP protocol stack as well as the relative position of the RTP/RTCP protocols.

Figure 7 - TCP/IP Protocol Stack

The user datagram protocol (UDP) is a best try, connectionless service. No prior connection needs to be established and packet overhead is lower than for connection-oriented services. RTP is using UDP to send packets. UDP is used since there is not usually enough time to transmit packets and lower packet overhead is desirable for real time traffic [10].

There is no error correction with UDP meaning that UDP has lower overhead and is more efficient for carrying real-time traffic. Furthermore, recovering packets that are in error, may cause conferencing traffic to lose its real time component which is essential. Consequently,

no advantage is gained in using the connection-oriented protocol TCP to carry RTP/RTCP PDUs.

The transmission control protocol (TCP) is a connection-oriented, reliable service. A TCP connection is established between two end points using a three-way handshaking protocol. The handshaking protocol exchanges sequence numbers and other information before a stream is transmitted. TCP establishes a full-duplex connection with sliding window flow control and go back N error correction. Real time traffic does not normally use TCP since the packet overhead is much larger than for UDP. TCP is used when errors cannot be tolerated and preventive measures such as three-way handshaking and error correction are required.

3.5 Multicasting Support

Multiple participant conferences are supported by the platform in order to test the synthetic stereo echo cancellation algorithm. IPv4 is not inherently capable of providing native multicasting abilities, which is the desired method of sending to multiple participants. Our platform needs to operate under the current IP standard, IPv4, therefore we considered two methods of multicasting support:

1. Use an Ethernet broadcast address to carry the data to all stations on the local LAN. This broadcast will only be allowed to pass through gateways, routers, switches and bridges if an mBone gateway server is present on the LAN.
2. Replicate packets at the host and direct them to multiple IP addresses

The first option is preferable since less load will be placed on the end terminals, since only one copy of the packet need be sent to all receiving participants in the conference. MBone is short for Multicast Backbone on the Internet. MBone is an extension to the Internet to support IP multicasting -- two-way transmission of data between multiple sites. The TCP/IP protocol used by the Internet divides messages into packets and sends each packet independently [33]. Packets can travel different routes to their destination which means that they can arrive in any order and with sizable delays between the first and last packets. In addition, each recipient of the data requires that separate packets be sent from the source to the destination. This works fine for static information, such as text and graphics, but it doesn't work well for real-time audio and video. With Mbone, a single packet can have multiple destinations and isn't split up until the last possible moment. This means that it can pass through several routers before it needs to be divided to reach its final destinations. This leads to much more efficient transmission and also ensures that packets reach multiple destinations at roughly the same time. The MBone is an experiment to upgrade the Internet to handle live multimedia messages. MBone servers have special Class D IP addresses [9][17][38].

The only drawback of this option is that it limits the number of conference calls to one per Ethernet. This is because only one broadcast address may be used in a LAN section at a time.

In the second option, the end terminals must replicate and send extra copies of each packet in a multiple participant conference. The load on the network will be lower with only one

packet being sent to a group of multiple recipients in method one making it desirable to network administrators.

3.6 Software Design

The real-time design of the VoIP software is illustrated through the use of use case maps. This design methodology shows the interaction between different concurrent segments of the system that are all performing in a real-time manner. The symbols and terminology used in this design methodology are shown in Appendix B.

A snapshot of the user interface of VoIP platform (dMIPS) end product may be seen in Figure 8. The reader may observe that the user interface of dMIPS is separated in sub-sections where different kind of information is either logged or being input. Host name and host IP address are automatically and display when program starts.

Figure 8 - Screenshot of dMIPS

The platform is research oriented and allows the user to experiment with several parameters (such as frame size, buffer size, thread priority and process class) in order to achieve optimum performance of the system. The user may also specify filename and path allowing him to keep acoustic round trip delay output file. The file contains sequence number and acoustic round trip delay values for each and every frame. The raw file has to be processed using simple matlab program in order to separate the two. The matlab file is given in Appendix A.

To start dMIPS the user specifies frame and buffer sizes and then simply presses "Join Conference". Then you will see "Play ISRs" and "Record TxISRs" numbers increasing. When a far end user joins the conference, Rx ISRs numbers starts increasing. As the conference progresses, RTP and RTCP local statistics are collected. These involve lost frames, frame count and interarrival jitter.

CHAPTER 4

4 Proposed method of Calculating round trip delay

This chapter proposes a theoretical model and an ITU-T H.323 compliant empirical measurement method for the acoustic round trip delay (ARTD) of a voice over Internet protocol (VoIP) conference. It also proposes a method for determining acoustic echo cancellation targets by combining acoustic round trip delay data and psycho-acoustic echo target data.

Accurate determination of echo targets is required to minimize detrimental switched-loss attenuation effects. A nonlinear processor (NLP) can use these echo cancellation targets to lower the amount of switched-loss attenuation added by the echo suppression system. Lower levels of switched-loss attenuation increase the system gain and reduce unwanted half-duplex effects where data (voice) is transmitted in only one direction at a time. Most VoIP conferencing products can use the proposed echo suppression system because it is H.323 compliant. The acoustic round trip delay is measured empirically by slicing the real-time signal path and identifying test points. On these test points, timestamps of each frame may be read and statistics may be updated. An acoustic domain verification test is used to check the ARTD empirical measurements and theoretical model for accuracy. A Carleton University VoIP platform is used to measure the ARTD.

4.1 Introduction

Voice over Internet protocol (VoIP) systems are of great interest to Public Service Telephone Networks (PSTN) companies. VoIP is a next generation technology that will capture some of the traditional circuit-switched network market share. The International Telecommunications Union – Telecommunications Sector (ITU-T) standard for VoIP is a toolkit standard which is called H.323 [21]. H.323 states the requirements for a visual telephone service over non-guaranteed quality of service networks, such as the Internet. It is important that any new methodologies introduced in this paper conform to H.323 as well the other relevant standards that it proposes. ITU-T H.245, [26], is used for capabilities exchange and for setting up ITU-T H.225.0, [24], logical channels. The H.225.0 logical channels use the Internet Engineering Task Force (IETF) real-time transport protocol (RTP), [19]. Figure 9, shows a high-level view of an H.323 terminal. The shaded components are involved in the acoustic round trip delay for a VoIP conference and were implemented in Carleton University DSP lab.

Figure 9 - High-Level View of H.323 Terminal Equipment

4.1.1 Testing Platform

A conferencing test platform was developed at Carleton University using a Digital Signal Processor (Tiger 542/PC DSP board or Tiger 548/PC DSP Board) and a personal computer (PC). A multimedia PC platform can also be used and it is common among typical users whereas the DSP platform involves a separate specialized DSP card. However, the DSP platform allows more control of the environment and the ability to run powerful signal processing algorithms that cannot be done with the multimedia PC platform. The multimedia PC platform must rely on the underlying operating system to service its events and is inherently less reliable than using a specialized piece of hardware such as the DSP card. It is expected that new developments in processor architecture, such as Intel's MMX technology, will continue to narrow the performance gap between these two platforms.

Current VoIP systems do not offer the means to measure or approximate the complete system round trip delay that the user perceives. This delay can also be labeled the acoustic round trip delay of the system since the user's perception of the VoIP conference delay is in the acoustic domain.

4.1.2 Quality of Service

The quality of service (QoS) of a VoIP system may be partly evaluated by the acoustic round trip delay. The level of talker confusion and difficulties experience in normal conversation protocol will be directly affected by this delay [1]. A conference system with

long delays causes users to engage in double talk and mutual silence more often. When delays increase, the likelihood of a participant interrupting the conversation increases. It is less likely that participants will respond to questions or pauses in the conversation [1]. These effects indicate an overall change in the normal handshaking protocol of a conversation. The effects may be considered as measures of the quality of service of the conversation as perceived of by the user.

4.1.3 Echo Suppression

Echo suppression is required for hands-free full-duplex conferencing. The echo suppressor prevents the loudspeaker signal from returning to the microphone transducer in the form of an acoustic echo. The coupling between the loudspeaker and microphone transducer is eliminated by the echo suppression system. This coupling is shown in Figure 10 as the reception room impulse response, $h(t)$. Practical implementations of echo suppression systems are composed of an acoustic echo canceller and a nonlinear processor (NLP) that add switched-loss [4][11].

4.1.4 Acoustic Echo Cancellation

The convergence criteria for an acoustic echo canceller is to minimize the error signal, $e(t)$, by adapting its model of the reception room, $\tilde{h}(t)$, to equal the reception room impulse response, $h(t)$ [33]. The error signal may be estimated by (4-1).

$$e(t) = d(t) - s(t) * \tilde{h}(t) = s(t) * [h(t) - \tilde{h}(t)] \quad (4-1)$$

Figure 10 –Acoustic Echo Cancellation Structure

Figure 10, shows a typical acoustic echo cancellation structure with reference signal, $s(t)$, the primary signal, $d(t)$ and the error signal, $e(t)$.

Echo return loss enhancement (ERLE) is a measure of the amount of echo power that is cancelled by the acoustic echo canceller. It is also a measure of acoustic echo canceller performance. Figure 10 shows the ERLE as it is measured across the adder in the acoustic echo cancellation structure. ERLE may be evaluated by the equation (4-2):

$$ERLE(db) = 10 \log \frac{E[d^2(n)]}{E[e^2(n)]} \quad (4-2)$$

Where, $d(n)$ is the primary signal and $e(n)$ is the error signal.

The maximum amount of ERLE that can be obtained for a specified number of taps is computed by the total impulse power to tail power ratio [12][28].

Figure 11 – Acoustic Echo Cancellation Structure with Switched-Loss

An acoustic echo canceller is unable to achieve perfect cancellation due to acoustic noise, quantization noise, truncation noise, undermodelling of the acoustic transfer function, enclosure vibration effects, transducer nonlinearities and algorithm tracking and convergence rates [12][33]. Due to these limitations, the maximum ERLE for an acoustic echo canceller is often insufficient for the conference. Practical implementations of echo suppression systems require the addition of switched-loss to overcome the acoustic echo canceller's limitations. Switched loss may be provided by a non-linear processor (NLP). Switched-loss reduces the overall system gain and adds unwanted half-duplex effects to the conference. Half-duplex effects can cause talker confusion, cut-off part of a user's speech due to switching delay and can allow a user to dominate the conference.

The nonlinear processor determines the level of switched-loss attenuation that needs to be added to the conference. Figure 11, shows an acoustic echo cancellation structure with a nonlinear processor that controls the addition of switched-loss. Four possible locations for the attenuators, L_1 , L_2 , L_3 and L_4 , that control the level of switched-loss in the conference [33], are also shown in Figure 11. Depending on system design requirements either all or some of these attenuators are present.

4.2 PsychoAcoustic Determination of echo target

Acoustic echo cancellation targets can be determined by combining acoustic round trip delay data and psycho acoustic echo target data. Studies show that the level of speech quality impairment due to echo is related to the volume of the received echo and the echo path delay. The amount of echo path loss is required for a given mean one way propagation delay is shown in Figure 12 [1]. These results are based on subjective testing of the psychological effects of echo on speech quality.

Figure 12 – Talker Echo Tolerance Curves

The TELR curves shown in Figure 12 are approximated by the following equation:

$$TELR = A_0 - 6e^{-0.3T^2} + 40 \log \frac{1+T/10}{1+T/150} \quad (4-3)$$

Where, $A_0 = 14\text{dB}$ for 1% objection or 7dB for 10% objection and T is the mean one way propagation time in ms.

The 1% objection curve is considered the acceptable case for network planing [23]. The 10% objection curve is only used in exceptional circumstances.

The talker echo loudness rating (TELR) curves are typically used by network planners to determine echo suppression targets for network echo cancellers. The echo loss for VoIP is the sum of the acoustic echo cancellation and switched loss obtained by the echo suppression system.

4.3 Theoretical model of acoustic round trip Delay

The components shown in Figure 13 affect the acoustic round trip delay for both the multimedia PC and DSP/PC platforms. The delays that we analyzed below are essentially the same for both platforms, except that the sound card platform requires an audio buffer with size 2 for play-out while the DSP platform requires a buffer with size 1 for play-out. This is because the DSP platform may service requests for new samples from the audio codec at the set the sampling rate since there is a devoted processor on the DSP board. On the other hand, on the PC platform a buffer with size 2 is required in order to have continuous play-out and avoid starvation of the system. The presence of a single processor in the PC platform makes the system slower. Sound card requests for playing out packets have to wait longer. The presence of an extra frame in the audio buffer will ensure that the packets will be played out continuously.

4.3.1 Explanation of Delays

Several assumptions concerning the nature of the two ends of the conference have been made in the process of developing of the proposed theoretical model. These assumptions are presented below:

Figure 13 – High Level Diagram for the Acoustic Round Trip Delay

1. All observations are made with reference to the start of a recorded audio frame.
2. Both ends are identical in hardware and software. (Same CPU power, operating system, memory configuration, software etc)
3. Recording, encapsulation and network delays are independent and they can occur simultaneously.
4. The delayed playback jitter buffer threshold is equal on both ends.
5. A new audio frame may be queued on the playback buffer concurrent with sound card or DSP playback.
6. Both participants are at steady state. It is assumed that the audio and jitter buffers are at their average values. The safest way to ensure that the participants are in a steady state, is to let the system run for few seconds before taking any measurements.
7. The room impulse responses for both ends are static and identical, therefore $t_a = t_a'$ i.e. the acoustic propagation delay at both ends of the conference is the same.

8. The sum of $(t_a + t_r)$ is constant for a station in steady state.

The delay components are detailed in the following table.

Table 1 Detailed Delay Explanation

t_f	The length of an audio frame in seconds.
Δt_p	Processing time taken to code the audio frame with the audio codec and to encapsulate the frame with RTP, UDP and IP. DSP to PC communication time.
Δt_n	Network one-way transmission and propagation delay.
Δt_r	Processing time to decapsulate the frame, calculate RTP statistics, generate an RTCP packet (if required) and the time to enqueue the audio frame in the micro buffer. IP, UDP and RTP decapsulation of received frame to obtain audio frame. Generate RTP statistics such as time and total received bytes. Generate RTCP packet if not indicated in incoming conference participants of pending network connection and if it is time to send. PC to DSP communication time.
Jitter Buffer Queue G/D/1	<ul style="list-style-type: none"> The jitter buffer size is JBS. The jitter buffer can be realized as a single server—single queue with General arrival and constant service time.
Audio Buffer D/D/1	<ul style="list-style-type: none"> The micro buffer size is ABS. The audio buffer can be realized as single server—single queue with Deterministic arrival and service distributions if it is deterministic assuming that the audio buffer is playing frames continuously (i.e. continuously). $JBS \leq \text{Frame length (in milliseconds) of the DSP to PC link}$.

Δp_i	<ul style="list-style-type: none"> Processing time taken to decode the audio frame with the audio codec
	<ul style="list-style-type: none"> Time taken for propagation time OR includes the propagation time that is the time taken for the signal to travel from the speaker to the microphone includes the delay time that is the time taken for the signal to travel from the microphone to the microphone Measured $t_{at} = d/v$ where d is the distance between the speaker and the microphone, v is the velocity of sound waves (330m/s) at room temperature OR measured by measuring the time delay of the microphone response to the impulse response of the room

The lack of synchronization between the record and the playback events requires compensation. This is achieved by introducing parameter t_r . This parameter represents the difference between the start of a recorded frame and the start of a playback frame since our reference in designing the model is the beginning of a frame. The lack of synchronization is caused by the random startup times for participants in the conference. The synchronization parameter is uniformly distributed between 0 and t_r . This implies that on the average this parameter is equal to $\frac{1}{2}t_r$.

$$0 \leq t_r \leq t_r \quad (4-4)$$

The self-synchronization parameter with acoustic delay, $t_r t_a$, adjusts the amount of time remaining in the recording frame during the timestamp loopback to account for acoustic

echo propagation. The complementary-synchronization parameter with acoustic delay, $t_s + t_a$, represents the offset time delay in the audio frame before the echo is played.

The lack of synchronization between stations in the conference requires the introduction of the cross-synchronization parameter t_v . This parameter represents the amount of time remaining in playback when a new frame is placed in the jitter buffer. It is distributed between 0 and t_f .

$$0 \leq t_v \leq t_f \quad (4-5)$$

A second set of delays that is derived from system architecture as described above is presented below on Table 2.

Table 2 Detailed explanation of Delays derived from system architecture

Delay	Explanation
$t_s + t_a$	<p>Self-Synchronization parameter with audio delay</p> <ul style="list-style-type: none"> • The amount of time remaining in the recording of an audio frame after a frame begins playback at the same station • Accounts for the lack of synchronization between record and playback events in a station • The empirical measurement model does not include the acoustic delay in the synchronization parameter • On the average, $t_s = 1/2 t_f$
t_v	Complementary Synchronization parameter with audio delay

Figure 14 – Pipeline analogy model of the ARTD path

The jitter buffer size, JBS_t , and the audio buffer size, $ABSt$, are values critical to system performance. If a new frame does not arrive during the time that the last queued frame spends in the jitter buffer, then the jitter buffer will “starve” by one frame. If the jitter buffer is allowed to starve by too many frames, eventually audible gaps will appear on the system playout. Then the system has to shut down the playout in order to re-buffer enough frames to start the playout again. Consequently, if $ABSt$ becomes zero then the conference system is not able to keep up with servicing the audio buffer of the sound card or codec of the DSP card, destroying the real time nature of the conference. Appropriate buffer sizing schemes for a frame voice receiver may be found in [3]. In our case, empirical observations were used in order to find the optimum buffer size for the best performance of our system.

The pipeline analogy starts with the first frame labeled, Frame 1, and proceeds with subsequent frames in increasing order. Each frame undergoes the same sequence of delays that were previously shown in Figure 13. A prior (quantitative) knowledge of the delay components and an accurate representation on the pipeline analogy figure may help a user to approximate the jitter buffer sizes of a platform.

4.4 Theoretical ARTD Equation

Mathematical relations can be derived through evaluation of Figure 14.

Event A: First frame queued:

$$T_A = t_f + \Delta p_{x1} + \Delta t_{n1} + \Delta p_{y1} \quad (4-6)$$

Event D: First frame played out:

$$T_D = t_f + \Delta p_{x1} + \Delta t_{n1} + \Delta p_{r1} + [JBS_i]_f + \Delta p_{z1} + ([ABS_i] - 1)t_f + t_v \quad (4-7)$$

Event Z: Total round trip delay for the first frame:

$$T_Z = t_f + 2\Delta p_{x1} + 2\Delta t_{n1} + 2\Delta p_{r1} + 2[JBS_i]_f + 2\Delta p_{z1} + 2([ABS_i] - 1)t_f + t_a + t_\gamma - t_a + t_f - t_\gamma + t_a + t_a + t_v + t_v \quad (4-8)$$

Consequently, Round Trip Delay is equal to:

$$ARTD = 2t_f + 2\Delta p_x + 2\Delta t_n + 2\Delta p_r + 2\Delta p_z + 2[JBS_i]_f + 2([ABS_i] - 1)t_f + 2t_a + t_v + t_v \quad (4-9)$$

Acoustic round trip delay empirical measurements that verify the above model are presented in the following chapter.

CHAPTER 5

5 RESULTS

5.1 Introduction

In this chapter, acoustic round trip delay measurements using the empirical method described in chapter 4 will be presented. Acoustic round trip delay may partly evaluate the quality of service (QoS) in a VoIP system (see 4.1.2). The empirical method used is H.323 and RTP compliant. The results obtained from the acoustic round trip delay empirical measurement method will verify the theoretical model. A verification model will be presented, and the empirical method results will be validated.

This chapter will also propose a method for psycho-acoustic determination of echo suppression targets. This method can be used to regulate the amount of switched-loss added to a hands-free full-duplex conference. The mean one-way propagation delay may be approximated by the acoustic round trip delay empirical measurement method proposed in section 3 of [36]. This method allows us to minimize switched-loss levels, which add detrimental half-duplex effects to the conference.

5.2 Empirical Measurement Method for Acoustic Round Trip Delay

The acoustic round trip delay is computed through the addition of timestamp information to the RTP header extension. This method is H.323 and RTP compliant since adding proprietary information to the header is a valid extension to the standard. Clients not supporting the acoustic round trip delay measurement service would ignore this header.

Figure 15 shows the progression of the timestamp information required to compute the round trip delay. The shaded blocks are included in the acoustic round trip delay empirical computation method. After recording has been completed, every frame is time-stamped at transmission time by the RTP layer of participant A. Thus, the initial recording delay, t_r , is not captured by the empirical method. This timestamp is signified as T_{old} and is extracted from station A. It is the time in milliseconds at the instant that was recorded.

Figure 15 – Empirical Measurement of Acoustic Round Trip Delay

When the frame reaches participant B, the timing information is looped back to the RTP transmission layer. This information is piggybacked onto the next frame that completes recording and is being encapsulated by RTP. The acoustic delay, t_a , is not captured by the timestamp loopback.

The effective playback audio buffer delay includes the cross-synchronization parameter, t_v . t_v is the amount of time remaining on the playback of an audio frame when an incoming

frame is placed in the buffer. The two stations are not synchronized (no synchronization source between the two) and a new frame send from station B to station A can arrive at any time during playback. Consequently, the effective audio buffer delay, EABD, may be defined as the amount of time an audio frame spent in the buffer and may be calculated by equation (5-1):

$$EABD = \{[ABS_t]\} t_f + t_v. \quad (5-1)$$

Where, ABS_t is the audio buffer size, t_f is the frame size in milliseconds, t_v is the cross synchronization parameter.

The reference for the empirical model is the beginning of a recorded frame, therefore, one frame must be subtracted from the $[ABS_t]$ since the audio information is at the very start of the frame.

After playback at station A the acoustic round trip delay is measured as the difference between the current system, $T_{current}$, and the old time contained in the RTP header extension, T_{old} . $T_{current}$ is the time in milliseconds at station A at the moment the frame is played out. The computation of the acoustic round trip delay is presented in (5-2).

$$ARTD_{Meas} = T_{current} - T_{old} \quad (5-2)$$

Referring to Figure 15, the amount of delay time that is computed from the above equation are the shaded areas in the figure. $ARTD_{Meas}$ may be equated to the actual delay parameters as equation (5-3) suggests:

$$ARTD_{Meas.} = 2[\Delta p_x + \Delta p_y + \Delta p_z + \Delta t_n] + 2\{[JBS_i]t_f + [ABS_i] - 1\}t_f + t_f + t_v + t_v' \quad (5-3)$$

The final acoustic delay, t_a' , as well as the acoustic delay at the far end (station B), t_a , are not included in the above empirical measurement method.

The empirical measurement method does not include the initial recording delay, t_f , or the acoustic delays for either participant. In order to get an accurate estimation of the acoustic round trip delay, a correction factor, C_F , must be introduced. C_F shall include the initial recording of a frame and the sum of the acoustic delays t_a and t_a' . As discussed in chapter 4, the sum of t_a and t_a' may be approximated as $2t_a$. Hence, C_F may be approximated as shown in equation (5-4).

$$C_F = t_f + 2t_a \quad (5-4)$$

Equation (5-5) suggests that the actual acoustic round trip delay is the sum of the measured round trip delay and the correction factor.

$$ARTD_{Corrected} = ARTD_{Measured} + C_F \quad (5-5)$$

Or, analytically:

$$ARTD_{Corrected} = 2[\Delta p_x + \Delta p_y + \Delta p + \Delta t_n] + 2t_f + 2t_a + 2([JBS_i] + [ABS_i] - 1)t_f + t_v + t_v' \quad (5-6)$$

Equation (5-6) is the same as the acoustic round trip delay for the theoretical model given in (4-9).

The dominant factors in the acoustic round trip delay are the audio frame size, the jitter buffer size, the playback buffer size and the voice coder delay. The user or the system designer sets the audio frame size, based on platform selection (MIPS or dMIPS), in order to optimize system's performance. The audio playback buffer delay is playback dependent. The multimedia PC platform requires a double audio buffer for playback that causes increased delay over the DSP/PC platform. The DSP/PC platform requires only a single buffer for playback. The voice coder delay can vary widely dependent on the coding scheme in use. A typical voice coder delay can be ranged from 20ms to 100ms [27].

The acoustic delay, t_a , is dependent on the room in use and the physical setup of the speakers and microphone. The acoustic delays are negligible compared with the network delays, jitter and audio buffer delays and voice coder delays. Therefore round trip delay may be approximated as the sum of the computed round trip delay, $ARTD_{Meas}$, and a correction factor of one frame, t_f .

$$ARTD = ARTD_{Measured} + t_f \quad (5-7)$$

5.2.1 Empirical Measurement method Results

Several measurements were taken using the empirical method on the DSP/PC platform. The results were taken for a local area network (LAN) VoIP conference.

Lost frames in the network may cause the ARTD computation method to give inaccurate results. A filtering method was developed in order to compensate this effect. This method computes an upper limit for the acoustic round trip delay based on prior knowledge of the system parameters. When a frame is lost in the network, the platform will ignore the round trip delay measurement for this frame and carry on the measurement for the next valid frame. Hence, lost packets in the network do not affect the system accuracy in a computation of the acoustic round trip delay.

The results shown in Figure 16, 17 and 18 verify (5-3). That is, the acoustic round trip delay computed by the system is equal to the sum of all delays presented in (5-3). Total delayed playback threshold (TDPT) is the total number of frames present in buffers (audio and jitter) at any instance after the system has reached steady state. TDPT may be calculated using equation (5-8). Each TDPT setting was ensemble averaged by performing five runs.

$$TDPT = [JBS] + [ABS] \quad (5-8)$$

Figure 16 - DSP Round Trip Delay Measurements for 20ms Frame

Figure 16, shows that when the number of frames in buffers increase, the round trip delay increases by a factor which is an integer multiple of a frame, t_f . For example, when the TDPT is equal to 6, this means that 6 audio frames are in buffers (round trip). Each frame is 20ms long. This gives us 120ms delay time in the buffers. Figure 16, shows that for TDPT equal to 6, the $\text{ARTD}_{\text{Meas}}$ is equal to approximately 140ms. The extra 20ms accommodate processing and network delays. When TDPT is increased to 8, we observe that the $\text{ARTD}_{\text{Meas}}$ is approximately equal to 180ms. Given that processing and network delays are constant, the only the difference between the two TDPT settings is about 40ms or two

frames. This proves that RDT in dMIPS is only dependent on the buffer sizes and independent of network delays.

Figure 17 - DSP Round Trip Delay Measurements for 30ms Frame.

Figures 17 and 18 are suggesting the same. In Figure 17, frame size is 30ms. It may be clearly seen that an increase of two frame in the total buffer size will affect the computed $ARTD_{Meas}$ by a factor of 2 frames or 60ms. Figure 18, suggests the same idea with frame size of 40ms.

Figure 18 - DSP Round Trip Delay Measurements for 40ms Frame

5.3 Acoustic Domain Verification of acoustic round trip delay

In order to verify the correctness and accuracy of the empirical results presented above, an acoustic domain verification method of the round trip delay was developed. The arrangement of Figure 19 was setup in the DSP laboratory at Carleton University.

Figure 19 - Verification Model

A digital audio tape (DAT) was connected to station A and 10 seconds of white Gaussian noise (WGN) was played in one of the 8 available tracks. At station B the speakers are coupled to the microphone. Back in station A, the "returned" WGN is recorded on a different track. Then, by cross correlating the two tracks, the difference (in samples) between the two tracks may be estimated. The cross correlation of white noise signals is an impulse function, since its power spectral density is a constant [2][7]. The relative position of the impulse function on the x-axis will help us calculating an exact result for actual the round trip delay (excluding the acoustic delays t_a and t_a').

Figure 20 – Cross Correlation of WGN for $t_f=20\text{ms}$

Figure 20 shows the cross correlation of the original signal and the delayed signal for frame size of 20ms. The impulse function is around 7,600 samples away from 0. By knowing the sampling rate of the white noise, the time difference between the two signals may be obtained. Figure 21 shows the cross correlation of the two white noise signals (original and delayed) for frame size of 40ms. The impulse function in this case is further away from the 0 point, at 12,400 samples, as we expect larger amount of delay.

Figure 21 – Cross Correlation of WGN Vectors for $t_f=40\text{ms}$

5.3.1 Calculations

Let the original WGN vector be $x(t)$ and the recorded, delayed vector be $y(t)$. A cross correlation of the two vectors may be obtained by:

$$R_{xy}(k) = \sum_{k=-\max lag}^{k=\max lag} x(t)y(t-k) \quad (5-9)$$

In order to calculate the delay between the two vectors, first, the position of the exact overlap of the two signals must be found (i.e. the impulse function). From Figure 20, it can be extracted that the impulse is located at 7,600 samples. Knowing that the noise was

sampled at 48,000 samples per second then the delay between the two vectors can be calculated:

$$ARTD_{Verified(20ms)} = \frac{7,600samples}{48,000samples/sec} = 158ms \quad (5-10)$$

This shows that the real round trip delay (excluding any acoustic delay) is 158ms. Similar calculations were also performed for $t_f=40ms$.

$$ARTD_{Verified(40ms)} = \frac{12,390samples}{48,000samples/sec} = 258ms \quad (5-11)$$

Comparing this to the empirical measurement for these settings the two methods proved to give the same results. In section 5.2, it was estimated that for TDPT equal to 6 and frame size of 20ms, $ARTD_{Meas}$ was equal to 140ms. After adding the correction factor, C_F , $ARTD_{Corrected}$ is equal to 160ms. The acoustic verification method gave us a result of 158ms. The two are very closed and the small difference may be attributed to small difference in processing delays. The result may be seen in Figure 22.

Figure 22 – Verification Results for $t_f=20\text{ms}$

Figure 23 compares results from the empirical method and the verification for 40ms frames.

Here, we may as well observe that the empirical method measured ARTD is about a frame ($t_f = 40\text{ms}$) lower than the verification results.

Figure 23 - Verification Results for $t_f=40\text{ms}$

The above measurements show that the empirical method results corrected with C_F are in a very close agreement with the acoustic method verification results. This proves the correctness and accuracy of the method proposed to measure acoustic round trip delay in VoIP systems.

5.4 Echo suppression target determination

A practical application of the acoustic round trip delay is the echo suppression target determination in a VoIP conference. Public Switched Telephone Networks (PSTN) do not have the ability to determine the acoustic round trip delay as perceived by an end user. However, VoIP conference systems that employ the acoustic round trip delay empirical measurement method have the ability to determine this delay.

Echo target can be determined by cross-referencing the mean one way propagation delay, approximated by one half of the acoustic round trip delay, with psycho-acoustic target echo loudness rating TERL targets. An example of echo target determination is shown in Figure 24. Psycho-acoustic data [22], [23] relates the level of TELR to the one way propagation delay.

Figure 24 – Psycho-Acoustic Determination of Switched-Loss

For a given TELR the amount of switched loss needed may be computed. This may be derived from equation (5-12).

$$TEL_R = ERLE_{AEC} + SL \quad (5-12)$$

Where, $ERLE_{AEC}$ is the maximum echo return loss enhancement for a specific echo canceller, and SL is the amount of switched loss required by the system.

The mean one way propagation delay allows the system to minimize the amount of switched-loss added to the conference. The amount of switched loss added to a conference should be minimized to avoid adverse half duplex effects. An example of the echo suppression system is depicted in Figure 24. Let mean one-way propagation delay be 100ms. For this delay the TERL is equal to 47dB and the amount of switched-loss, SL , is equal to 7dB. If the maximum mean one way propagation delay was 300ms, then 14dB of switched loss would need to be added to the conference.

Summarizing, Figure 25 combines ARTD results and echo tolerance curves. Using this figure, we will try to estimate how much switched loss, SL , is required based on acoustic round trip delay results. The echo target can be determined if the mean one way propagation delay (approximately one half of ARTD) is known. Frame sizes of 20ms and 40ms were used for this illustration. For 20ms frames the mean one-way delay was found to be 80ms while for 40ms frame was found to be 153ms. Projecting this information onto the Psycho-Acoustic curve, the exact switched loss required can be found. Assuming that the maximum $ERLE$ for an acoustic canceller at steady state is 40dB (as shown in the figure) then the required switched loss must be 4.6dB and 10.5dB for the 20ms and 40ms frames, respectively. It is obvious that the longer the frame is, the greater the mean one way propagation delay is, hence, more switched loss, EL_{SL} , is required.

Figure 25 – Combined ARTD and Echo Tolerance Curves

5.5 Conclusions

In this chapter, acoustic round trip delay was measured using an empirical method. Acoustic round trip delay may evaluate the quality of service (QoS) in a VoIP system. The empirical method used is H.323 and RTP compliant. A theoretical model was first developed. The results obtained from the acoustic round trip delay empirical measurement method verify the theoretical model. A verification model was developed, and the empirical method results were validated. This chapter also proposed a method for psycho-acoustic determination of echo suppression targets. This method can be used to regulate the amount of switched-loss added to a hands-free full-duplex conference. The mean one-way propagation delay must be approximated by the acoustic round trip delay empirical measurement method proposed in [36]. This method allows us to minimize switched-loss

levels, which add detrimental half-duplex effects to the conference. PSTN are unable to measure the delay and use excessive amounts of switched-loss which make hands-free full-duplex conferencing over digital networks a preferable option.

CHAPTER 6

6 CONCLUSIONS – FUTURE WORK

This thesis proposed a method for calculating the acoustic round trip delay in Voice over IP conferences. It also proposed a method for psycho acoustic determination of echo targets using results obtained from the acoustic round trip delay calculation. The acoustic round trip delay can be also used to partly evaluate the quality of service in VoIP conferences.

The acoustic round trip delay was measured through an empirical method that is H.323 compatible. The real time protocol (RTP), which is part of the H.323 suite, was used in the development of an acoustic round trip delay empirical measurement method. RTP is a transport protocol providing end to end services for data with real time characteristics such as interactive audio and video. A theoretical model was presented and empirical results collected verify the theoretical model. An acoustic domain verification method was also used to ensure that results from the empirical measurement were accurate. The results were then used to determine the desired ERLE target from psycho-acoustic data. ERLE targets can be used by the NLP to minimize the level of switched-loss injected into a conference.

In the above process, two VoIP conference platforms were developed: The DSP/PC platform called dMIPS, and the Multimedia/PC platform called MIPS. Both conform to the ITU-T H.323 standard. The development of these two platforms was essential since there were no commercially available systems that could be customized for VoIP research.

6.1 Future Research

The DSP/PC platform, dMIPS, is crucial to further research that will be performed in the area of VoIP at Carleton University in terms of signal processing and network studies. dMIPS contains a high power DSP board that is capable of running complex signal processing algorithms that future researchers may propose. The multimedia PC platform, MIPS, can be ported to an MMX PC for signal processing with MMX processor enhancements. Both platforms are highly configurable and can be used to test performance under various network conditions or modified to test new networking algorithms or features. Appropriate buffer sizing may be calculated using methods proposed in [3] in order to reduce latency in packet voice receivers.

The impact of IPv6 and the reservation protocol (RSVP), [29], on VoIP system design could be examined. The combination of these protocols could reduce network latency and enhance the quality of service (QoS) for VoIP conferences. Furthermore, bandwidth management schemes for VoIP systems employing RSVP may be studied.

Implications of the operating system and system drivers on VoIP system performance issues could be also further studied. A statistical model for various parameters of the system could be developed based on a theoretical model. The model may be tested using simulation or by modifying dMIPS and MIPS platforms.

7 References

- [1] P. T. Brady, "Effects of Transmission Delay on Conversational Behavior on Echo-Free Telephone Circuits," *Bell Labs Technical Journal*, pp. 115-134, v. 50, 1971.
- [2] H. Stark, J. W. Woods, Probability, Random Process and Estimation Theory for Engineers, Prentice Hall, Englewood Cliffs, NJ, 1979.
- [3] G. Barberis, "Buffer sizing of a Packet Voice Receiver," *IEEE Transactions on Communications*, pp. 152 – 156, vol. Com-29, No 2, February 1981.
- [4] B. Widrow and S. Streams, Adaptive Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1985.
- [5] D.D. Clark and D.L. Tennehouse, "Architectural considerations for a new generation of protocols," *SGCOMM Symposium on Communications Architectures and Protocols*, Philadelphia, Pennsylvania, pp. 200-208 IEEE Sept. 1990. *Computer Communications Review*, Vol. 20(4), Sept. 1990.
- [6] S. Haykin, Adaptive Filter Theory, 2nd ed., Prentice Hall, Englewood Cliffs, N.J., 1991.
- [7] D. Bartsekas and R. Gallanger, Data Networks, Second edition, Prentice Hall, NJ, 1992.
- [8] Mills, D., "Network Time Protocol Version 3", RFC 1305, UDEL, March 1992.H.
- [9] Y. Rekhter and T. Li, "An Architecture for IP address allocation with CIDR," RFC 1518, September 1993.
- [10] K. Jeffray, Stone D.L and Smith F. D, "Transport and display mechanisms for multimedia conferencing across packet-switched networks," *Computer Networks and ISDN systems*, pp. 1281-1304, vol. 26, 1994.

- [11] M. E. Frerkin, Digital Signal Processing in Communication Systems, Van Nostrand Reinhold, NY, 1994.
- [12] M. Knappe and R. Goubran, "Steady State Performance Limitations of Full-Band Acoustic Echo Cancellers," Proc. ICASSP94, pp. 2073-2077, April 1994.
- [13] K. Egevang, P. Francis, "The IP Network Address Translator (NAT)," RFC 1631, May 1994.
- [14] W. Simpson, "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, Daydreamer, July 1994.
- [15] J. C. Bolot, T. Turletti, and I. Wakeman, "Scalable feedback control of multicast for video distribution in the internet," SIGCOMM Symposium on Communication Architectures and Protocols, London England, pp. 58-67, ACM, Aug. 1994
- [16] J. Reynolds, and J. Postel, "Assigned numbers", STD 2, RFC 1700, USC/Information Sciences Institute, September 1994.
- [17] R. Atkinson, "Security Architecture for the Internet Protocol," RFC1825, August 1995.
- [18] M. Sondhi, D. Morgan and J. Hall, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," IEEE Signal Processing Letters, vol. 2, no. 8, August 1995.
- [19] Schulzrinne, S. Casner, R. Frederick and V. Jacobson, IETF Audio-Video Transport Working Group RFC 1889, "RTP: A Transport Protocol for Real-Time Applications," Jan. 1996.
- [20] R. Atkinson "IP Authentication Header," RFC 1826, April 1996.

- [21] ITU-T H. Series, "H. 323 – Line Transmission of Non-Telephone Signals - Visual Telephone Systems and Equipment for Local Area Networks which Provide a Non-Guaranteed Quality of Service," May 28, 1996.
- [22] ETSI Technical Report, "ETR 250 –Transmission and Multiplexing TM; Speech Communication Quality from Mouth to Ear for 3,1 KHz Handset Telephony Across networks," July 1996.
- [23] ITU-T Series G., "G. 131 –Transmission Systems and Media – International Telephone Connections and Circuits – General Characteristics of the 4- wire Chain Formed by the International Circuits and National Extension Circuits – Control of Talker Echo," Aug. 1996.
- [24] ITU-T H. Series, "H. 225 – Line Transmission of Non-Telephone Signals – Media Stream Packetization and Synchronization on Non-Guaranteed Quality of Service LANs," Dec. 26, 1996.
- [25] G. Thom, "H.323: The Multimedia Communications Standard for Local Area Networks," IEEE Communications Magazine, pp. 52-56, Dec. 1996.
- [26] ITU-T H. Series, "H. 245 – Line Transmission of Non-Telephone Signals – Control Protocol for Multimedia Communication," March 24, 1997.
- [27] R. Sudhir G. Ahuja A. Krishna and T. Muti, "Packet Telephony," Bell Technical Journal, Spring 97.
- [28] N. Brikett, "Nonlinear Adaptive Filtering with Application to Acoustic Echo Cancellation," Ph.D Thesis, Carleton University, Ottawa, Canada, April 1997.

- [29] A. Mankin, F. Baker, B. Braden, S. Braden, and M. Romanow, A. Weinrib and L. Zhang, "Resource ReSerVation Protocol (RSVP): Applicability Statement Some guidelines on Deployment," RFC 2208 September 1997.
- [30] W. Stallings, Data and Computer Communications, Fifth Edition, Prentice Hall, Englewood Cliffs, NJ, 1997.
- [31] T. Yensen, R. Goubran and I. Lambadaris, "A Multiple Participant CTI Acoustic Echo Cancellation Structure," IEEE NORSIG98, pp. 193-196, June 1998.
- [32] R. Braden, D. Clark and S. Shenker, "Integrated Services in the Internet Architecture: An Overview," RFC 1633, September 1998.
- [33] T. Yensen "Structures and Interfaces for Voice Over IP," Master Thesis, Carleton University, Ottawa, Canada, September 1998.
- [34] M. Biegi, R. Jennings, S. Rao and D Verma, "Supporting Service Level Agreements using Differentiated Services," draft, November 1998.
- [35] K. Nichols, S. Blake, F. Baker and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," RFC 2475, December 1998.
- [36] T. Yensen, M. Parperis, I. Lambadaris, and R. Goubran, "Determining Acoustic Round Trip Delay for VoIP Conferences," IEEE, MMSP98, December 1998.
- [37] M. Crawford, "Transmission of IPv6 Packets over Ethernet Networks," RFC 2464, December 1998.
- [38] Internet Engineering Task Force (IETF) Mbone Working Group: "IP Multicast Applications: Challenges and Solutions," May 1999.
- [39] <http://www.mbone.com>
- [40] <http://www.newbridge.com>

[41] <http://www.ngi.gov>

[42] <http://www.canet2.net>

[43] <http://www.playground.sun.com>

APPENDIX A

Software Design

The VoIP software design is illustrated with the help of *use case maps*. *Use case maps* show the interaction between concurrent system segments that execute in real-time. The symbols and terminology, used in use case maps are shown in Figure 26.

Figure 26 - Use Case Map Symbols

A-1 High-Level Use Case Map

The high-level use case map, Figure 27, shows the major system blocks which are broken down further into more detailed use case maps. These blocks reside in the DSP

card domain or the PC domain. The dashed lines in the use case maps indicate the distinction between these two domains.

Figure 27 – High-Level Use Case Map for RTP/DSP Software

A-2 Jitter Buffer Use Case Map

The jitter buffer use case map, Figure 28, shows the copying of data from the host port interface shared memory to the DSP jitter buffer on receipt of a PC jitter buffer interrupt.

Figure 28 - Jitter Buffer Use Case Map

A-3 Jitter Buffer Activity Descriptions

Activity	Description	Activity Performed by
Copy Data from Shared Rx Buffer to Jitter Buffer	This action occurs when the PC interrupts the DSP to indicate that incoming data for a participant is resident in the host port interface shared rx buffer. The DSP must copy this memory to the appropriate participant's jitter buffer.	DSP Card Software

Table 3 - Jitter Buffer Activity Descriptions

A-4 Send Data Use Case Map

The send data use case map, Figure 29, shows transmission of a captured audio frame through encapsulation by RTP and sending over UDP/IP.

Figure 29 - Send Data Use Case Maps

A-4-1 Send Data Activity Descriptions

Activity	Description	Activity Performed by
Receive Data from DSP Card via PIO	Data is transferred from the DSP shared card capture buffer to the PC using the programmed I/O transfer mode. (PIO)	PC Software
Add RTP Header	Data received from the DSP card is encapsulated with a valid RTP header to form an RTP frame.	PC Software
Send UDP Datagram	The RTP frame is sent to the remote system(s) via the UDP/IP using IP multicasting.	PC Software

Table 4 - Send Data Activity Descriptions

A-5 Receive Data Use Case Map

The receive data use case map, Figure 30, shows the decapsulation of incoming RTP packets, validation of RTP packets, passing audio frames to the DSP and generation of RTCP statistics.

Figure 30 - Receive Data Use Case Map

A-5-1 Receive Data Activity Descriptions

Activity	Description	Activity Performed by
Receive UDP Datagram	An RTP frame is received from the network via the UDP protocol with IP multicast.	PC Software

Activity	Description	Activity Performed by
Validate Frame and Collect Stats	The payload type is verified and the sequence number is checked to determine whether the packet is current or expired. The jitter is calculated, the total packet count is accumulated, and other statistics are computed.	PC Software
Remove RTP Header	The RTP frame header is verified and the data is removed from the RTP structure.	PC Software
Delete Frame	The RTP data that has been receive is non-current or corrupt and is deleted.	PC Software
Add to Jitter Buffer, Delete RTP Frame	The RTP data that has been removed from the frame is copied to the shared rx buffer in the host port interface memory. The DSP software will be responsible for adding it to the correct jitter buffer for the participant. The RTP frame is then deleted.	PC Software
Send RTCP Frame	Send an RTCP Sender Report (SR) frame to the designated participant when the percentage bandwidth is exceeded.	PC Software

Table 5 - Receive Data Activity Descriptions

A-6 Multiple Participant Jitter Buffer Structure

The multiple participant jitter buffer consists of space to store n participant's incoming audio frames up to a maximum depth of 10 frames per participant. The jitter buffer operates on a delayed playout methodology whereby network jitter is smoothed by buffering up several frames before playout begins. The number of frames buffered before playout is indicated by Δ in Figure 31. The participant's activity status and IP address must also be logged by the jitter buffer for playback of data and synchronization with the appropriate spatialization function.

Jitter Buffer Structure

Figure 31 - Jitter Buffer Structure

A-7 Host Port Interface and Transfer Registers

The audio data frames and control information is transferred from the PC to the DSP and vice versa through the host port interface (HPI), shared memory. The memory allocations within the DSP are shown in Figure 32. The transfer registers and their addresses in PC I/O space and DSP I/O space are also shown.

Figure 32 - Host Port Interface and Transfer Registers

A-8 Capture and Jitter Buffer

The capture and jitter buffer host port interface memory allocations are shown in Figure 33.

Round Trip Delay Computation Information + Optional Components (Host Port Interface Memory)

		<u>Constant Name</u>
0x143E	Jitter Buffer Size	JBUFSIZE
0x143F	Jitter Buffer Restarts	JRESTARTS
0x1440	PC Cur. Timestamp, $T_s(n)$	PCURTS
0x1442	PC Old Timestamp, $T_s(n-1)$	POLDTS
0x1444	PC Current Seq #, $S_n(n)$	PCURSN
0x1445	PC Old Seq #, $S_n(n-1)$	POLDSEN
0x1446	Round Trip Delay	RTD
0x1448	Sequence # Delta	DSN
0x1449	DSP Old Seq. #, $S_n(n-1)$	DOLDSEN
0x144A	DSP Cur. Timestamp, $T_s(n)$	DCURTS
0x144C	DSP Old Timestamp, $T_s(n-1)$	DOLDTS
0x144E	Unused[946]	BLOCK
0x17FF		

Figure 34 - Round Trip Delay Information

A-10 Typical Control Use Case Map

A typical control use case map is shown in Figure 35. This use case map scenario is followed whenever control information is changed by the PC and sent to the DSP. The DSP must acknowledge that the control information has been received.

Control Use Case Map

(Startup Example)

Figure 35 - Typical Control Use Case Map

A-11 Matlab program to read ARTD and sequence number.

```

function [rtd, sn, dsn] = readrtd(filename)

% function [rtd, sn, dsn] = readrtd(filename)

% Read Round Trip Delay data from RTP_VOIP

%

% Post:

%   RTD - Round Trip Delay
%   SN  - Sequence Number
%   ASN - Change in Sequence #

%

% Note: Round Trip Delay data file is composed of round trip delay
% information in uint32 and sequence number in uint16. Therefore
% special interlaced file reading must be used.

% By: Marios Parperis

% Read Round Trip Delays

rtd = fread(fid,inf,'uint32',2);

% Read Sequence Numbers

frewind(fid); %Rewind to read sequence numbers

fread(fid,1,'uint32');    %Dummy read

```



```
sn = fread(fid,inf,'uint16',4);
```

```
for n=2:length(sn)
```

```
    dsn(n-1)=sn(n)-sn(n-1);
```

```
end
```

```
fclose(fid);
```

APPENDIX B

Hardware/Software requirements for dMIPS platform

- Pentium Class PC System or Greater
- Windows NT Workstation 4.0 or Windows '95/98
- Visual C++ v5.0 32-bit (Development Only)
- PCI 10Mb/s Ethernet Card (Twisted Pair 10BaseT)
- Winsock 2.0 for Windows NT 4.0 Workstation or Windows '95 (IPv4)
- Driver X Software
- Stereo Multimedia Speakers with Internal Amplifier
- Monaural Microphone
- Microphone Pre-Amp with Line Level Output to DSP Card
- DSP's Code Composer Software (Load DSP Code)
- Tiger 542/PC DSP Board or Tiger 548/PC DSP Board Based on the TMS320C542 or TMS320C548 respectively.

Jumper settings for the Tiger 542/PC DSP board or Tiger 548/PC DSP Board Based on the TMS320C542 or TMS320C548 respectively are:

- OVL off (connect pins 2 and 3),
- TBCOFF no jumper,

- ALN no jumper,
- ROM, no jumper,
- IRQ select IRQ 10,
- SI select IO port 2A0h ON-OFF-ON,
- MIC 2Vrms JP16 and JP18 jumper installed,
- OGAIN not set.