Investigating Factors Influencing QoS of Internet Phone

Difu Su, Jaideep Srivastava
Dept. of CS&E University of Minnesota
http://www.cs.umn.edu/Research/mmdbms
{sdf, srivasta}@cs.umn.edu

Jey-Hsin Yao
Fujitsu Labs of America

jyao@fla.fujitsu.com

Abstract

An increasing number of applications are using the Internet for voice transmission, and hence there is a growing demand for the Internet phone. However, the Internet was not originally designed for such isochronous traffic, and it is not clear that it is fully capable of providing the desired Quality of Service (QoS) for this application. In this paper we describe a number of QoS parameters to measure the perceptual quality of audio, including measures of delay, rate, and loss. We investigate the factors influencing the QoS of Internet voice, such as packet period and packet size, threshold delay, buffer delay, time slots and Internet sites. We present results from our experiments on how these factors influence the QoS of Internet voice. Based on the results of our experiments, we propose a QoS driven dynamic scheduling algorithm for the real-time transmission of Internet voice packets.

1. Introduction

With the increasing popularity of the Internet in people's daily life, continuous media (*CM*) services on the Internet play an increasingly important role. Internet voice or Voice on Networks (*VON*) is a new and important CM service. The applications of VON can be divided into non-real-time and real-time. For real-time applications, voice delay is critical. We cannot rely on a large buffer and reliability through transmission, which result in long delay, to provide good QoS. The Internet phone is a real-time application, and has advantages over the regular phone service. It can reduce the expense for long distance calls, reach new domains of users, and provide a platform for value-added applications[1]. There are over 30 audio/video phone products providing the VON service today, with its popularity on the rise. However, because the quality of VON is not high enough, "Internet telephony is still an early adopter market." [2]

QoS parameters for the Internet phone include delay, rate variation, and loss in the quality of voice transmission. Since the Internet phone is interactive, only small amounts of voice delay, rate variation, and loss rate are acceptable [9]. Voice is transmitted in packets on the Internet. This may result in a long delay and a high loss rate. Given the nature of the phone service, among the three types of QoS parameters, voice delay is the most crucial. Many factors influence the QoS of the Internet phone. Hence, it is useful to determine which factors affect the QoS of the Internet, and how. These results can be used to schedule playback in the Internet phone to get the desired voice quality. They are also helpful for the pricing of the Internet telephony[3].

In the next section, we provide some background on the Internet phone. In section 3 we present our investigation of the factors that influence the QoS of Internet phone. In section 4, we propose a dynamic playback scheduling of the Internet voice to provide good QoS for Internet phone. Finally, we present our conclusions and future directions in section 5.

2. Background

In this section we first discuss some protocols for transmitting voice on the Internet. We then present our QoS metrics for measuring the perceptual quality of voice.

2.1 Protocols for transmitting voice

Phone voice is usually sampled at the rate of 8,000 Hz. If we use one byte to encode a sample, then the data rate for the phone voice is 8,000 bytes (*B*) per second.

If we pack the set of samples in time ΔT into a packet, the relationship of the size of a voice packet and ΔT is given by Equation 1. Since every ΔT time a new voice packet is generated, we call ΔT the *period* of the voice packets in this paper. This may result in a longer network delay and loss rate. Some common values for ΔT are 20ms, 40ms, and

0-7695-0253-9/99 $10.00 © 1999 IEEE

60ms, and the corresponding sizes of voice packets are 160B, 320B and 480B.

$$\boxed{Packet\ Size = 8000 \times \Delta T}$$

Equation 1

The Internet Protocol (*IP*) is the underlying protocol of the Internet, but is usually not used directly by applications. Transmission Control Protocol (*TCP*) and Uniform Datagram Protocol (*UDP*) are often used by applications instead. Given the limitations on network bandwidth, a straightforward TCP implementation is not suitable for real-time interactive applications, as shown by Smith [4].

UDP is a connectionless, unreliable protocol, which lacks the facilities for supporting real-time services. Hence it is not used directly for real-time services. Instead we use the Real-time Transport Protocol (*RTP*) [6] for the end-to-end delivery services for data with real-time characteristics.

RTP carries data that has real-time properties, and relies on the RTP Control Protocol (*RTCP*) to monitor its QoS. Although RTP itself does not ensure timely delivery or provide other QoS guarantees, its services make it easy for the application to manage its QoS requirements. In our implementation, RTP runs on top of UDP, and hence all the features and services of UDP are inherited.

Our RTP packet consists of a fixed RTP header and the payload. The fixed RTP header includes a sequence number, payload type, timestamp, and some other fields. We rely on the sequence number and timestamp to compute the delay and reorder the packets. The payload data is the voice samples. More details about RTP can be found in [7].

2.2 QoS metrics for internet phone

Internet phone aims at providing a voice quality that is comparable (or better) than the regular phone service, at a lower cost. Since the service is for use by people, it is natural to consider metrics of audio perceptual quality to measure the quality of the service provided. Wijesekera [9] defines a suite of metrics that do consider lossy CM streams, and are thus more suitable for measuring the transmission of CM streams on RTP. The measurable quality parameters of the Internet phone include voice delay, rate variation, and voice loss. These measures reflect the continuity of a voice stream, and can be mapped into the *rate variation, timing drift* and *content loss* in Wijesekera's model, which we adopted for our study.

A voice stream can be envisioned as a continuous flow of voice samples, referred to as logical data units (LDU) in the uniform framework of [8]. Given the ideal rate and the beginning time of a CM stream, there is an ideal arrival time for each LDU. The actual arrival time of a LDU may deviate from the ideal time. The rate variation may be measured more accurately by the drift parameters. The aggregate drift factor (*ADF*) specifies the aggregate drifts from the ideal over a number of consecutive LDU's in a stream. The consecutive drift factor (*CDF*) specifies the maximum consecutive non-zero drifts from the ideal over a number of consecutive LDU's in a stream.

The loss of content of a CM stream may be measured by two factors. The aggregate loss factor (*ALF*) is the aggregate number of LDUs lost over some time interval in a stream, and actually it specifies the percentage of loss in a stream. The consecutive loss factor (*CLF*) is the maximum number of consecutive LDUs lost in a stream, and is directly related to the longest time during which voice is lost. An example for explaining the definitions of these four QoS metrics is shown in Figure 1. A stream of 10 LDUs is shown in Figure 1. The *ith* LDU is expect to arrive at time *i*. The top line shows the actual arrival time of each LDU. The 2nd, 3rd, 5th and 10th LDUs are late. Their jitters are 0.2, 0.2, 0.3, and 0.3 respectively. ADF is 1.0 or 10% of the total time. CDF is 0.4, the sum of the 2nd and 3rd jitters. The 6th, 7th and 9th LDUs are lost. The ALF is 30%. The CLF is 2 LDUs.

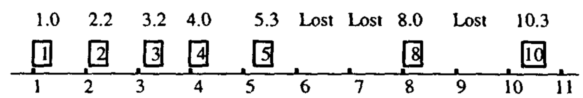


Figure 1 Examples for ALF, CLF, ADF and CDF

In [10] results of a detailed perceptual study for audio and video are described, where the impact of various QoS metrics described in [9] was studied. For an audio stream, up to 21% of aggregate loss and approximately 2 consecutive LDU loss is acceptable. Further, up to 7% of audio rate variations is tolerable. The upper bound for CDF is not given. These results are important for understanding the tradeoffs between loss and rate variation, to provide an acceptable overall QoS of Internet voice.

3. Study of QoS of Internet voice

In this section, we present our study on the factors that influence the QoS of Internet voice. We first investigate the factors influencing voice quality. Then we present our experiments and results.

3.1 Factors influencing voice quality

The quality of Internet voice transmission may be influenced many factors, including the transmission protocols, the status of the Internet, time slots, the period and size of voice packets, locations of the two communicating sites, and the playback scheduling. The relationships among delay, jitter, and loss of RTP packets are complex.

$$\boxed{\begin{aligned} Voice\ Delay = {} & Generation\ Delay + \\ & TransmissionDelay + Buffering\ Delay \end{aligned}}$$

Equation 2

Let's begin with analyzing the delay of Internet voice. The delay of the Internet voice consists of three parts shown in Equation 2. The *generation delay* is the time required for sampling the voice and generating the voice packet. The generation delay is fixed for a given packet period. In this paper, we simply use packet period as a rough estimate of the generation delay. The *transmission delay* is the time for the RTP packets to reach the destination. The transmission delay depends on the status of the Internet. The *buffer delay* is the time the voice packet is stored in the buffer before it is replayed. According to Equation 2, a short buffer delay helps to reduce the total voice delay. However, if the buffer delay is too short, we have to play the voice soon after it arrives. This may result in too much jitter in voice. Silence has to be inserted before the late packets, and the last part of the late packet has to be truncated. In order to reduce voice jitter, an appropriate buffer delay is preferred.

Given a *maximum acceptable delay* as the *threshold delay*, a packet with longer delay than the threshold is considered lost. It is reasonable that if other conditions remain unchanged, a larger threshold results in a lower loss rate, a longer average delay and higher jitter.

We call the period of the voice packet the *packet period.* We refer to the size of the voice packet when we use the phrase *packet size* in this paper. The packet size is decided by the packet period according to Equation 1.

Because the generation delay is roughly equal to the packet period, we know that a short packet period helps to reduce the voice delay from Equation 2. We also know that a packet with a shorter period has a smaller packet size from Equation 1. However, within a fixed period of time, a shorter period results in more packets and thus more network traffic. This may result in a higher loss rate, longer delay and more jitter.

The traffic load of the Internet and the load of the OS vary from over time. This may influence the delay, loss and jitter of RTP packets. In high traffic load, these QoS parameters tend to reach unacceptable levels.

The transmission delay is influenced by the status of both the LAN and the WAN. Instinctively, transmission delay varies with the network distance between the two communicating computers. The longer the distance, the longer the transmission delay.

3.2 Experimental design & implementation

We design experiments to investigate how the factors described above influence the QoS of Internet voice. The idea is to collect the timing data for the delay and loss of RTP packets, and to analyze these data. To collect the data, a process (the *Sender*) keeps sending RTP packets to the process (the *Receiver*) in another computer.

The model of the experiments is shown in Figure 2. The receiver is located at the receiver computer and the sender is located at the sender computer. The two computers are connected via the Internet. The sender keeps sending RTP packets to the receiver. The receiver stores the data for timing and loss in files for later analyses. The *time checker* process in the receiver computer periodically sends packets to the *time server* to check the time difference between the two computers.

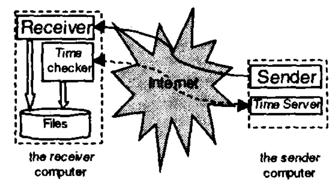


Figure 2 Model of the experiments

We implement the RTP specified in [7] on top of the UDP services provided by Java. The RTP services that we used in our experiments include payload type identification, sequence numbering, and timestamping.

3.3 Results and Analyses

The QoS parameters for Internet voice include delays, ALF, CLF, ADF, and CDF as stated in 2.2. From a statistical point of view, for delays we consider *mean delay*, *minimum delay*, *maximum delay* and *standard deviation* of delay. The *mean delay* is the minimal delay for the voice to be replayed at the destination. The *minimum delay* is the delay of the earliest voice packet that is available at the receiver. The *maximum delay* is the delay of the latest voice packet that can be replayed. The difference between the minimum delay and the maximum delay is the minimum buffer size required for the voice. The *standard deviation* is a measure of the voice jitter. A smaller value of the standard deviation means less jitter.

In our experiments, ALF is the loss rate, which means the percentage of loss for Internet voice. CLF is the longest time without voice. ADF is a measure of the drift of the voice rate, which means the percentage of rate variation. CDF is the longest consecutive nonzero drift time. Though CDF is a measure of the drift, it has no direct meaning in our experiments, so we decide not to include CDF in our analysis.

3.3.1 Effect of threshold delay

In these experiments, the sender computer is *vermouth.ee.umanitoba.ca* and the receiver computer is *rawana.cs.umn.edu*. *Vermouth* is a SUN UltraSparc 1 workstation running Solaris V2.6, and it is located at the University of Manitoba in Canada. *Rawana* is a SUN UltraSparc 1 workstation running Solaris V2.5, located at the University of Minnesota in the U.S.A. The LAN connected to each computer is a 10M bps Ethernet. The experiments lasted for 4 hours on October 24^{th}, 1998, which is a Saturday. The packet period is 20 ms and the packet size is 160 bytes. The buffer delay is 0. The results are shown in Figure 3 (a), (b), (c), and (d).

Figure 3a shows the influence of threshold on delays. The threshold does not influence the minimum delay, and has a week influence on the mean delay. The maximum delay is always close to the threshold value, so the buffer size is decided by the threshold value. Figure 3b shows the standard deviation of the delays. The standard deviations are very small meaning that most delays are close to the mean delay. Figure 3c shows the ALF and ADF. Only 1.3% of voice has a delay between 70ms and 200ms and most delays are under 70ms. Figure 3d shows CLF decreasing from 140ms to 80ms as the thresholds increase from 70ms to 200ms. CLF is insensitive to the change of threshold.

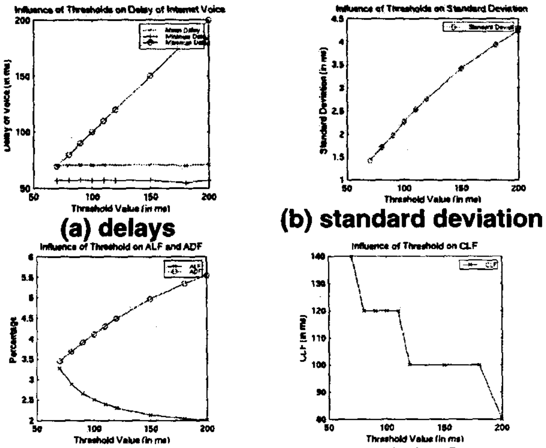


Figure 3 Effect of threshold value

3.3.2 Effect of buffer delay

The only interesting effect of buffer delay is its influence on ADF. We use the data collected from the experiments in 3.3.1. The results are shown in Figure 4.

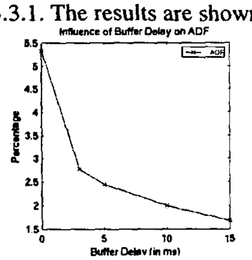


Figure 4 Effect of buffer delay

When the buffer delay increases from 0ms to 3ms, ALF drops sharply from 5.3% to 2.75%. When the buffer delay increases from 3ms to 15 ms, ADF decreases from 2.75% to 1.7%. When the buffer delay is longer than 5ms, the influence of its variance on ADF is not so strong. Compared with threshold's influence on ADF, buffer delay's influence on ADF is much stronger.

3.3.3 Effect of packet period

In these experiments, the sender computer is Vermouth and the receiver computer is Rawana. The packet period decides the packet size according to Equation 1, so when we talk about packet period in this subsection, we also refer the relative packet size. We varied the packet period while keeping other conditions unchanged. Each experiment lasted for 2 hours on October 25[th] 1998, a Sunday. The threshold delay is 180ms. The buffer delay is 0. The results are shown in Figure 5(a), (b), (c), and (d).

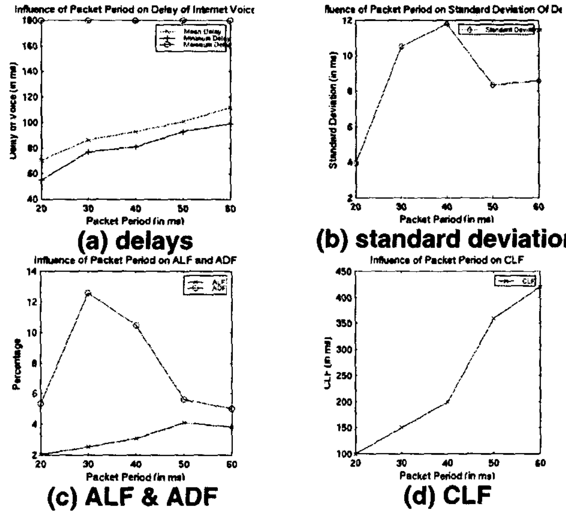


Figure 5 Effect of packet period

Figure 5a shows the influence of packet periods on delays. According to Equation 2, we get the transmission delay. We find that the RTP packets of different packet sizes have nearly the same mean transmission delay and the same minimum transmission delay. The effect of the variance of packet period on the mean and minimum transmission delay is negligible.

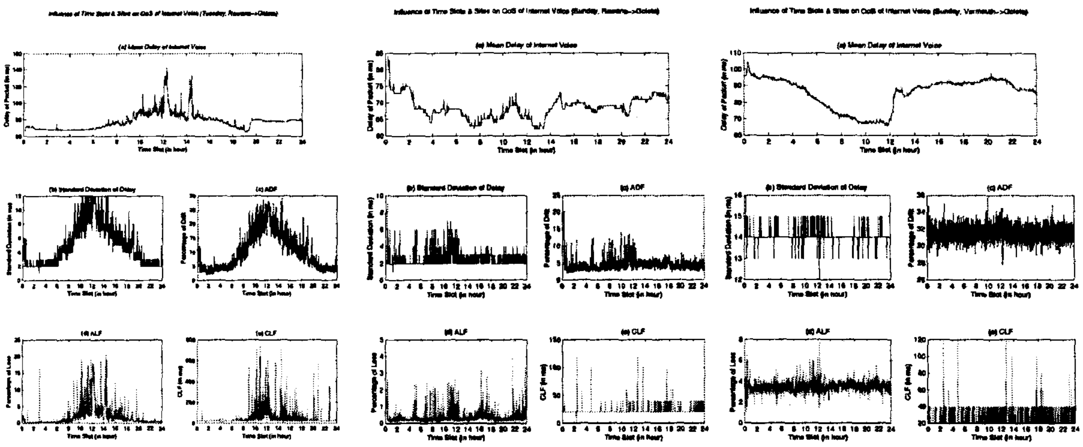
Figure 5b shows the standard deviation of the delays. The effect of the packet periods on the standard deviation is unclear.

Figure 5c shows the ALF and ADF. ADF reaches its peak at a packet period of 30ms. At 20ms, 50ms, and 60ms ADF has approximately equal values. We believe the exceptions of ADF at packet period of 30ms and 40ms are due to different time slots. This seems to imply that the effect of the packet period's variance on ADF is not strong. As the packet period increase from 20ms to 60ms, ALF increases from 2% to 4%. A longer packet period results in higher ALF. Since a longer period results in a larger packet size, but more packets during the same amount of time, we conclude that the effect of a larger packet size is stronger than the effect of fewer packets.

Figure 5d shows CLF. A longer packet period results in longer duration of voice loss.

3.3.4 Effect of time slot and site

In these experiments, the sender computers are *Vermouth.ee.umanitoba.ca* and *Rawana.cs.umn.edu*. The receiver computer is *Goleta.fla.fujitsu.com*. Goleta is a SUN UltraSparc 1 workstation running Solaris V2.6 located



(a) (b) (c)

Figure 6 Effect of time slots and sites

at Sunnyvale in California, connected to a LAN with a 10Mbps Ethernet. Each experiment lasted for 24 hours. The threshold value is set dynamically, with a value of 20 ms more than the mean delay. The packet period is 20 ms and the packet size is 160 bytes. The buffer delay is 0. To investigate the influence of time slots, we ran the experiments on a Sunday and a Tuesday from Rawana to Goleta. To investigate the influence of sites, we ran another experiment from Vermouth to Goleta on a Sunday. We compute the QoS parameters over the data for every minute for each experiment. The results are shown in Figure 6 (a), (b) and (c).

First, investigate the effect of time slot. Figure 6a shows the statistical data of Tuesday from Rawana to Goleta. Figure 6b shows the statistical data of Sunday from Rawana to Goleta. All the curves show dynamic features. In most of the time slots of a day, the mean delay curves have a value between 60ms to 80ms, except at some peaks. A peak lasts for less than 10 minutes and most of the peaks last for only 1 or 2 minutes. At peaks, the mean delay curves jump to a value between 100ms and 150 ms. The peaks appear mainly at daytime (roughly between 9am and 6pm) of weekdays. During the peaks, all other QoS parameters become worse: the larger standard deviation of delays indicating more jitter; and much higher ADF, ALF and CLF meaning more jitter and loss of voice.

During weekend and weekday nights, most of the ADF's are less than 7%, except at a few peaks where the values may reach 12%. During daytime of weekdays, the curve goes up until noon with a top value of 28% and goes down in the afternoon. Peaks with much higher values happen frequently. The ALF and CLF curves have the same tendency as the ADF curve.

Overall, all QoS metrics of Internet voice show dynamic behavior. We can divide the time slots of a week into two types, the *high quality slots (HQS)* and the *low quality slots* (*LQS*). The time slots of weekend and nights(roughly from 6pm to 9am) belong to HQS, and other time slots belong to LQS. An HQS shows less dynamic behavior than a LQS. An HQS has a shorter mean delay, less jitter, and a lower loss rate, so the QoS of Internet phone is easier to guarantee. Because both HQS and LQS show dynamic behavior in all the QoS parameters, dynamic scheduling for the playback of Internet voice is required to provide good QoS.

Figure 6c shows the statistical data of Sunday from Vermouth to Goleta. Comparing Figure 6b with Figure 6c, we find that the values of all the QoS parameters are different. However, they both show dynamic behavior on all the QoS parameters. Furthermore, since both experiments were run on a Sunday, they both have fewer and lower peaks in all the curves than that in Figure 6a. The effects of time slots of different sites are similar.

3.3.5 Conclusions from the experiments

From the above experiments, we draw the following conclusions.

- A shorter packet period results in better QoS, and 20ms is preferred according to our experiments.
- Increasing buffer delay helps to reduce ADF. A buffer delay of 3-5ms is preferred. When buffer delay is more than 15ms-20ms, the effect of its variance on ADF is not efficient.
- All QoS parameters show dynamic behavior. The time slots of a week can be divided into HQS and LQS. Dynamic scheduling of voice playback is required to provide good QoS.
- The mean delay of the voice packets can be used as the minimal voice delay to replay Internet voice. However, mean delay shows a dynamic behavior, and hence the receiver should dynamically compute the mean delay to get a good estimation on the future mean delay.

- The threshold value can be used to adjust the ADF, ALF, and CLF. A long threshold can be used to reduce ALF and CLF, but raise ADF. However, when the threshold is 20ms longer than the mean delay, the effects of its variance on ADF and ALF are not so pronounced. So, usually a threshold of lower than 20ms plus the mean delay is preferred.

4. Dynamic playback scheduling

Based on the results of our experiments, we propose a QoS driven scheduling algorithm for the playback of Internet voice. Our objective is to minimize packet delay while keeping other QoS parameters within acceptable levels. The methods that can be used to adjust the delay, ADF, ALF, and CLF have been discussed in 3.3.5. In our scheduling algorithm, we assume that the packet period is 20ms (packet size is 160B), and we have enough buffer for voice. We also assume that the user specifies the acceptability thresholds for ADF, ALF and CLF.

We use a queue to keep the timing information of the last minute to compute the mean delay, ADF, ALF and CLF. At startup, we set the buffer delay to 3ms and the threshold to be the mean delay plus 20ms. When a voice packet arrives, we put its delay into the queue. If a packet has not arrived within its threshold delay, we mark it lost and use the threshold delay as its delay. We put the loss mark and the threshold delay into the queue. We remove the oldest information if necessary, and compute the new mean delay, ADF, ALF, and CLF.

Next, we adjust the threshold value and buffer delay according to other QoS parameters, and add the same amount of change as that of mean delay to the threshold delay value. If the ALF and the CLF are higher than the acceptable level, we increase the threshold value by 1ms if it is less than 200ms. This helps to reduce the loss of voice but may increase jitter and slightly increase the mean delay. If the ALF and the CLF are lower than the acceptable level, we decrease the threshold value by 1ms if it is more than 20ms. This helps to reduce jitter and the mean delay but increases ALF and CLF. If ADF is higher than the acceptable level, we increase the buffer delay by 0.5ms if it is less than 20ms. This helps to reduce ADF in the subsequent voice packets but may increase the mean delay. If the ADF is lower than the acceptable level, we decrease the buffer delay by 0.5ms if it is more than 3ms. This helps to reduce to mean delay but increase the ADF. After these adjustments, we are ready to use the new buffer delay and the mean delay as the delay to play the voice thereafter. We will use the new threshold delay to judge if a packet is lost.

The key point of our playback scheduling is the estimation of the future voice delay. Our QoS driven scheduling use the information of the last minute as estimation. Using this scheduling, we can get a short delay while other QoS parameters are also kept at an acceptable level if possible.

5. CONCLUSION AND FURTHER WORK

In this paper, we investigate some factors that may influence the QoS of the Internet phone. These factors include packet period and packet size, threshold delay, buffer delay, time slots and Internet sites. We present our experiments on how these factors influence the QoS of Internet voice. Based on our experimental results, we proposed a QoS driven dynamic playback scheduling to provide good QoS for the Internet phone. We expect this scheduling to provide short voice delay and acceptable overall QoS. Due to the time constraint, we haven't implemented the scheduling. The further works include implementing the scheduling and evaluating its performance.

Thanks to Nobuaki Kawato and Tomohiko Taniguchi of the Fujitsu Labs of America for their help.

6. REFERENCES

[1] Kathy Meier. "The vision. Slides of Spring VON conference," April 2, 1997. Available at site http://www.von.org.

[2] Blake Irving. "Why is the talking net so quite?" Slides of Spring VON conference, April 2, 1997. Available at site http://www.von.org.

[3] Takeo Hamada, Jey-Hsin Yao, Difu Su. "Reflection on Internet and Telephony Integration –from the pricing point of view-," Internet Routing Quality of Service, Proceedings of SPIE, Vol. 3529. Nov. 1998.

[4] Brian C. Smith. "Cyclic-UDP: A Priority-Driven Best-Effort Protocol," 1994. Available at site http://www.cs.cornell.edu/Project/zeno/Papers/cudp.pdf.

[5] Zhigang Chen, See-Moong Tan, Roy H. Campbell, and Yongcheng Li. "Real Time Video and Audio in the World Wide Web," Fourth International World Wide Web Conference, Boston, Massachusetts, December 1995.

[6] H. Schulzrinne. "Issues in designing a transport for audio and video conferences and other multipariticipant real-time applications," Available at site ftp://gaia.cs.umass.edu/pub/hgschulz/rtp/issues.ps, Oct. 1993.

[7] H. Schulzrinne. "RTP: A Transport Protocol for Real-Time Applications," Available at site ftp://ftp.isi.edu/internet-drafts/drafts-itef-avt-rtp-new-01.txt, August 7, 1998.

[8] Steinmetz, Ralf and Blakowski, Gerold. "Media Synchronization Survey: Reference Model, Specification and Case Studies," IEEE Journal on Selected Areas in Communication. Volume 14, 1996. Page 5-35.

[9] Duminda Wijesekera and Jaideep Srivastava. "Quality of Service (QoS) Metrics for Continuous Media. Multimedia Tools and Applications," Volume 3, September 1996. Page 127-166.

[10] Duminda Wijesekera, Jaideep, Srivastava, A. Nerode, and M. Foresti. Experimental "Evaluation of Loss Perception in Continuous Media," ACM-Springer Verlag, 1998.