A Framework for Adaptive Voice Communication over Wireless Channels

Sandeep K. S. Gupta and Suhaib A. Obeidat
Department of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287
{Sandeep.Gupta, Obeidat}@asu.edu

***Abstract*— The varying channel error conditions of a wireless link do not allow for providing a steady service quality. Adaptability is one way to go around this problem since the quality of service (QoS) requirements can be changed as the channel status changes. In this paper we propose a framework for adaptive voice over one-hop wireless communication between a source and a destination. Adaptability is achieved at two levels: (i) changing the voice-encoding rate, and (ii) changing the signal modulation scheme. In addition to reacting to the channel conditions and hence achieving better quality, adaptation will allow for the support of more of the offered load (i.e., higher multiplexing gain). This will introduce some issues and problems, the goal was to reach to a near-optimal operating point, where quality is not traded for bandwidth efficiency. We simulated a special case of the framework. Results show that the adaptive approach has a much less degradation in voice quality (DVQ) and allows for more voice communications to be supported.**

***Index Terms*— modulation, multimedia communications, packetized voice, wireless communications.**

I. INTRODUCTION

During the past several years, there has been an interest in packetized voice, since it will allow for the convergence and integration of data, voice, and video traffic. Integration will expectedly result in lower costs by allowing for one access point and same network infrastructure for all sorts of traffic. Voice, as a real-time application, is sensitive to loss, delay and delay variations (jitter) with different degrees. The interactivity and real-time nature of voice require strict QoS guarantees from the network. Long delays affect the interactivity of voice, where cross talking starts to occur, a phenomenon that we observe when using Internet Telephony. For the same reason, losses cannot be handled by retransmissions as in the case of data traffic, simply because this will result in long delays. Additionly, since voice is an isochronous application, it is sensitive to delay jitter [1][2][3]. For these reasons, voice is usually transmitted as constant rate traffic, with no losses or delays (e.g., over TDMA). Even though this provides the quality needed, it is very costly. Since voice is inherently variable (human voice has a speech activity of about 42% [3]); the uncompressed (i.e., constant-rate) voice does not take advantage of the silence periods, and hence doesn't achieve any multiplexing gain. The support of real-time applications on wireless networks has recently become a hot area. Wireless networks are error-prone; the wireless channel's status is dynamically changing as the signal is exposed to absorption, scattering, interference, and multi-path fading. These two issues, namely 1) the changes in the wireless channel's conditions, and 2) the variability nature of voice traffic, can both be addressed by an adaptive scheme in which 1) sources do not send at a fixed rate (hence, we can achieve multiplexing gain), and 2) when the channel is in bad condition, we can send less-compressed traffic. This paper presents a framework of this adaptive scheme over a one-hop wireless communication. Two parameters of adaptation are introduced. The encoding rate, and the signal modulation scheme. When the channel is in good condition, sources send at a lower rate (16 kbps), and dense modulation is used (QAM16). On the other hand, when the channel is in deep fading, no compression is involved (i.e, sources send at 64 kbps), and BPSK modulation is used. We simulated a special case of the framework to get an insight about its performance.

A. Motivation

Case 1: To motivate the work, let's consider the following scenario: Given that the wirelss link has a 256 kbps capacity, if the link is in bad state, every source will be sending uncompressed voice at a rate of 64 kbps. The maximum number of sources that can be accommodated will be 4. However, if the channel state is good, each source sends compressed voice at a rate of 8 kbps. The same link now can accommodate 28 sources sending at 8 kbps. This resembles the case where all sources are in the same state (i.e., all good or all bad). Table 1 shows the different combinations where different sources have different channel condition.

As we can see, to allow for worst-case scenario, and using only adaptive encoding, no adaptation can be introduced. However, this can be changed by introducing FEC and adaptive modulation. If FEC is used, sources can still send compressed voice even if the channel state is bad. Portion of the bandwidth savings resulting from compression can be used for error control.

TABLE I

NUMBER OF SOURCES THAT CAN BE ACCOMMODATED WHEN USING ADAPTIVE ENCODING.

No of Sources(Bad State)	No. of Sources(Good State)	Total Supported
4	0	4
3	8	11
2	16	18
1	24	25
0	30	30

TABLE II

NUMBER OF SOURCES THAT CAN BE ACCOMMODATED WHEN USING ADAPTIVE MODULATION.

No of Sources(Bad State)	No. of Sources(Good State)	Total Supported
4	0	4
3	4	7
2	8	10
1	12	13
0	16	16

Case 2: Here, let's assume the wireless link has a capacity of 256 symbols/sec. And no adaptive encoding is used. When channel state is bad, BPSK modulation is used, link capacity is 256 kbps (similar to case 1 above). When channel is good, and using QAM 16, link capacity is $256 * 4$ kbps. Again, this resembles the case where all sources are in the same state. Table 2 shows the different combinations where different sources have different channel condition.

II. ADAPTABILITY

A. Adaptability Aspects Considered

As we mentioned, uncompressed voice does not take advantage of the silence periods. In addition, the dynamics of the wireless channel introduce variance in the amount of losses. Modulation schemes allow for sending more data using the same bandwidth. This can be achieved by encoding more bits per symbol. The denser the modulation is, the more the number of bits per symbol. However, this is not free of charge; the more the number of bits per symbol, the higher is the probability of bit errors. This is due to the inability of the receiver to detect the signal with an accuracy to distinguish the very small changes in the signal (whether in phase, frequency, etc.). As a result, the denser the modulation is, the more the number of bit errors (i.e., BER will be higher). Figure 1 below illustrates these ideas. Adaptive modulation proved to provide better results than non-adaptive modulation. In this work, we will consider adapting both the voice encoding level, and the modulation scheme used. When the channel is in a good state, denser modulation and highly compressed voice are used. On the other hand, when the channel is in deep fading, and the number of losses is high, we use sparse modulation and less-compressed voice. This way, when the channel is in a good condition, we can achieve high multiplexing gain, with a good quality. And when the channel is in a bad condition, we try to maintain a reasonable quality, which means less multiplexing.

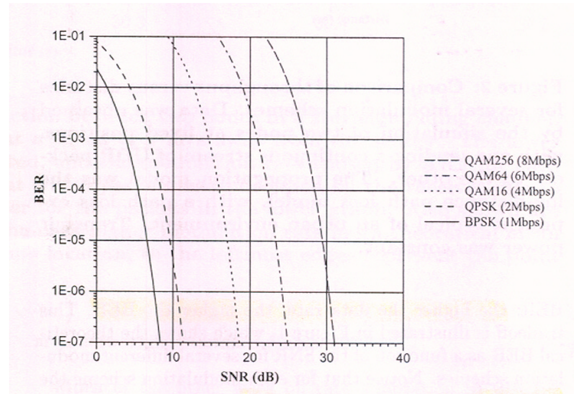

Fig. 1. SNR vs. BER for several modulation schemes [4].

B. Previous Work

Many variable-rate schemes for multimedia traffic have proposed in the literature. In [5], a variable channel coding scheme for Adaptive Multi Rate (AMR) speech transmission is proposed, where the ratio of voice bits and error control bits adapts depending on the wireless channel condition. Shenker [6] compared strict versus adaptive applications, showing that rate-adaptive applications react to network congestion better than other adaptive classes (e.g., delay-adaptive applications that keep sending at the same rate but change their delay requirements). Meo et al [7] studied the behavior of rate-adaptive voice communications over IP networks, results show that the adaptive approach allows for more voice communications while maintaining an acceptable quality. Adaptive modulation in wireless networks has also been addressed in the literature [4][8][9][10]. Reacting to the channel state by changing the modulation scheme and symbol-rate-used has proved to provide better performance. This has motivated newer wireless devices to support different modulation schemes [4].

III. THE SIMULATION MODEL

In this section we present the simulation model. This includes the framework under which adaptive voice is quantified, the voice source model, and the model for the wireless channel.

A. Framework

Figure 2 below shows the framework under which we studied and analyzed the effectiveness of adapting compressed voice sources across a wireless link. Uncompressed voice (64 kbps) is fed to an encoder that reduces the number of bits needed to represent the voice signal. We assume that the encoder is capable of coding voice to match a target size . The number of bits used to encode voice affects the quality of the compressed voice. Issues such as algorithmic delay, computational complexity and robustness to background noise are beyond the scope of this work. After the output bit stream has been encoded, voice signal is modulated using a particular modulation scheme based on the current status of the wireless channel. If the channel state

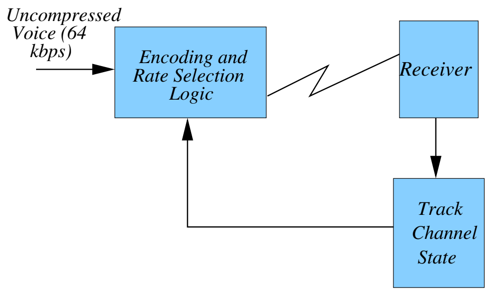

Fig. 2. General framework for adaptive voice over a wireless channel

is good, denser modulation is used. On the other hand, if the channel state is bad, a sparse modulation is used. The voice signal is then injected into the wirelss link. The channel status is collected at the receiver and sent back to the sender. Furthermore, current wireless devices have more than one modulation scheme, in addition to the logic needed to switch from one scheme to another. This means that they support different data rates [4].

In the above figure, the voice sources are multiplexed at an access point that transmits the aggregate traffic over the wireless channel. The link bandwidth is 1.544 Mbps (T1 rate). In traditional TDMA, a T1 rate can accommodate 24 voice channels. In our case, we will vary the number of voice sources and experiment on how many sources can be accommodated.

B. Voice Traffic Model

Modeling of voice traffic has been studied extensively in the literature, which resulted in several models. For a detailed explanation of voice models see [12]. In general, these models represent speech as a Markov Chain with different number of states. The more the number of states in the model, the more complicated it is. Among the simplest models for voice traffic is the On-Off model first proposed in [13], where a two-state chain is assumed. This is the most widely used model for its simplicity, and highly acceptable results. The two states correspond to the talk spurt and silence periods. Figure 3 shows the model and an example of its event sequence [12]. As shown in the event sequence of the Figure, the activities are exclusive, i.e. while A is talking B is silent (ATBS), and vise versa (ASBT). This means that the model doesn't account for double talk, and mutual silence. Hence, it doesn't model two-way conversations accurately. In the figure, p is the probability that A stops talking and B starts. Similarly, r is the probability that B stops talking and A starts. The holding time in the talk state and the silence state is assumed to be exponentially distributed with mean times of 352 ms and 650 ms, respectively [14]. These values are highly acceptable in the published literature.

C. Generic Description of the Scheme

As we mentioned earlier, the fluctuating BERs of a wireless channel motivate adaptation at the sources in a way that the quality delivered at the destination is not highly varying and hence more stable. In this subsection we will present a generic description of our scheme in a multi-state channel, and in the next subsection, we will present a 2-state model that is used in the experiments. Fluctuations in the BER of a wireless channel are modeled by a finite-state Markov Chain [15]. Figure 4 shows a Rayleigh channel of k states. The states resemble different Signal to Noise Ratio (SNR) ranges. A State i resembles an SNR that falls in the range $[S_i, S_{i+1}]$. State 0 has the lowest SNR (hence, highest BER) and state $k-1$ has the highest SNR (lowest BER). Since state $k-1$ resembles the best channel state, we can take advantage of that by using highly-compressed voice, and highly dense modulation. Similarly, when in state 0, we do not use compression at all, and we use the most sparse modulation. Rate selection will be proportional to the change in the BER. For example if state A has a BER = X_1 with an encoding rate of R_1 bits/sec, and the channel then moves to state B that has a BER = $2X_1$ (double of the previous state), then the encoding rate that will be selected is $2R_1$ (i.e., compression is *reduced* to half of the previous state). This way, we are maintaining a stable QoS (in terms of losses) that is as good as the worst channel condition. When the channel is in better situation, we are achieving the same quality but with less resources. Selection of modulation scheme is done in a similar fashion. It is assumed that the channel can move from one state only to one of its neighboring states. Since a Rayleigh model has an exponential pdf, the average time spent in a particular state i is exponentially distributed and is equal to 1 / $(\alpha_i + \beta_i)$.

D. Wireless Channel Model

We are modeling the wireless channel using a two-state Elliot-Gilbert model [16] [17]. The two states represent the good (G) and bad (B) states, see Figure 5. Each state has an

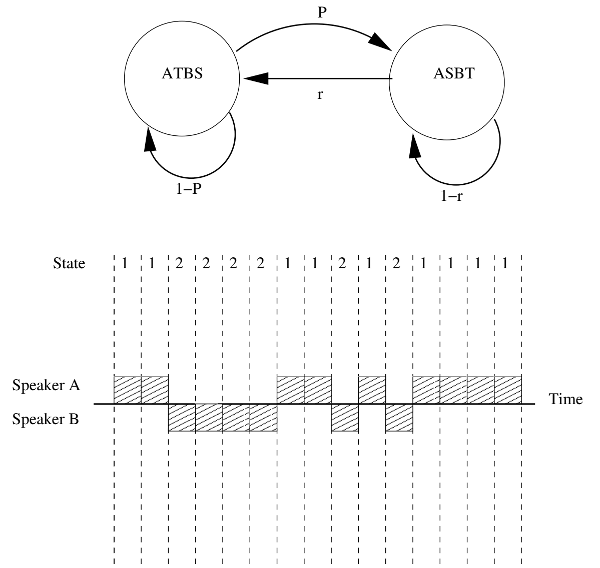

Fig. 3. Voice as a two-state Markov Chain, with the associated event sequence

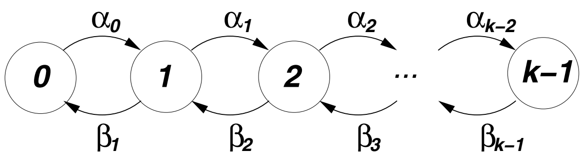

Fig. 4. Generic Model of the Wireless Channel

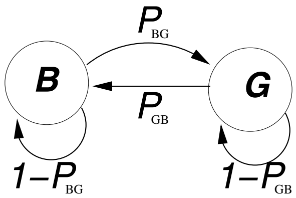

Fig. 5. Gilbert-Elliot Channel Model

associated error probability. $Pe(G)$ is the error rate in the good state and $Pe(B)$ is the error rate in the bad state. In the figure, P_{BG} and P_{GB} represent the transition probability of going from bad to good, and good to bad states, respectively. In our experiments, when the channel is in the good state, voice sources transmit compressed voice at a rate of 16 kbps, and when the channel is in the bad state, they send at a rate of 64 kbps. The justification for this is that losing a compressed voice packet has higher cost than losing a non-compressed one. As a result, when the channel is in bad condition, we want to avoid costly losses. On the other hand, when the channel is in a good condition, we can send compressed voice, and achieve higher multiplexing. We also change the modulation scheme used. When the channel is in good state, we take advantage from denser modulation, in this case we used QAM16, and when the channel is in bad condition, BPSK is used.

E. Simulation Parameters

All simulations were run with a confidence interval of 95% and an accuracy of 10%. Each simulation was run 25 times to achieve the confidence interval and accuracy mentioned above. Number of voice sources considered 24,40 ,55,70 etc. Wireless channel's bandwidth: 1.544 Mbps (T1 rate). The packet size is fixed and equal to 32 bytes. This value is reasonable for voice communications, as the larger the packet size, the longer the packetization/depacketization delay (especially that we are considering compression). We ran the simulation for 120 sec. Senders follow the two-state model, with mean values of 352 ms, and 650 ms, for the talking and silence periods, respectively. The two channel states have an average residency time of: 10 sec in the good state and 4 seconds in the bad state. The BER of the good state: $Pe(G) = 10^{-6}$, and that of the bad state $Pe(B) = 10^{-2}$. The size of the buffer at the access point has a size of 200 packets.

F. Results and Analysis

For the simulation study, we considered two cases (i) Adaptive Modulation versus non-adaptive modulation where adaptive encoding is used in both of them, and (ii) Adaptive Encoding versus non-adaptive encoding where adaptive modulation is used in both. The reason for this is that we wanted to see the effect of each factor on the quality achieved separately.Figure 6 shows packet loss ratio when using the adaptive modulation versus the non-adaptive one (case 1 above). In both cases, adaptive encoding is considered. This was in order to see how much effect adaptive modulation alone has on the total number of losses. As we can see from the figure, for number of sources less than 85, the adaptive approach has a higher packet loss ratio. After that, the it starts to show better peformance. The next few paragraphs explain the reason behind this.

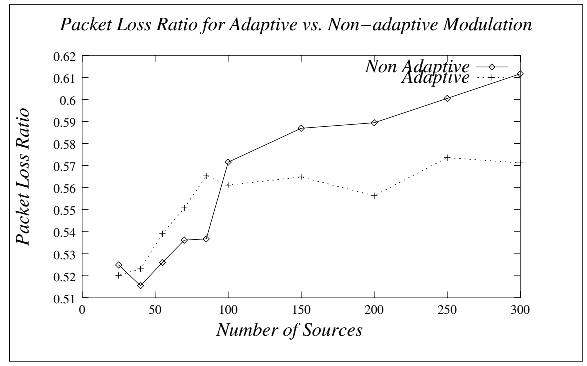

Fig. 6. Packet Loss Ratio for Adaptive vs. Non-adaptive Modulation

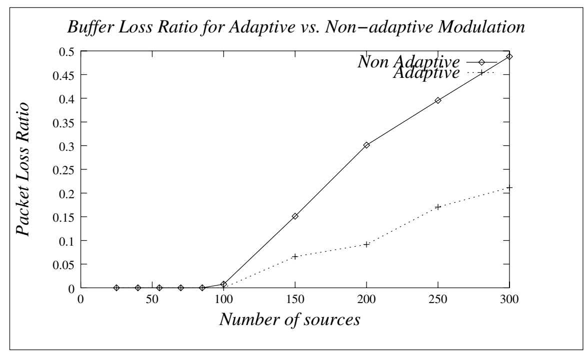

Fig. 7. Buffer Loss Ratio for Adaptive vs. Non-adaptive Modulation

In order to understand this behavior, we need to look into the behavior of loss components separately. The components of loss are (i) buffer losses, and (ii) channel losses. Figures 7 and 8 show buffer and channel losses, respectively. Initially, buffer losses are similar for adaptive and non-adaptive modulation and both are equal to 0. This resembles the case where all traffic generated from the voice sources is being handled instantly by the channel and no queuing is needed or queue is never full. As the number of sources starts to increase, we will start to notice the difference between the two schemes. As can be seen from Figure 7, buffer losses are constantly less for adaptive modulation than for the non-adaptive one. Adaptive modulation has a loss ratio that is less than half of the non-adaptive one.

Since the adaptive scheme serves more packets (i.e., less

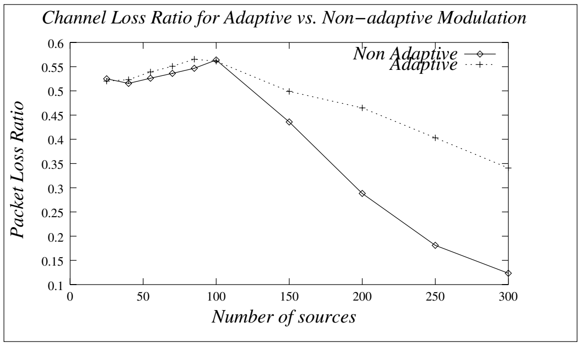

Fig. 8. Number of Channel Losses for Adaptive vs. Non-adaptive Modulation

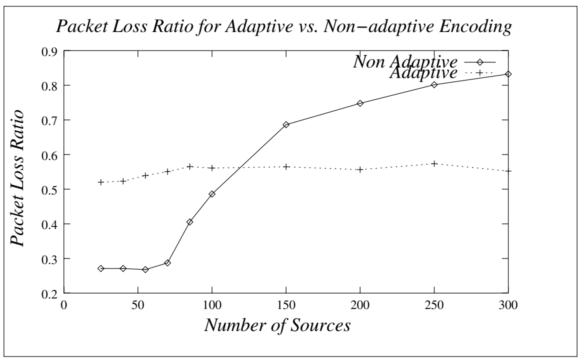

Fig. 9. Packet Loss Ratio for Adaptive and Non-adaptive Encoding

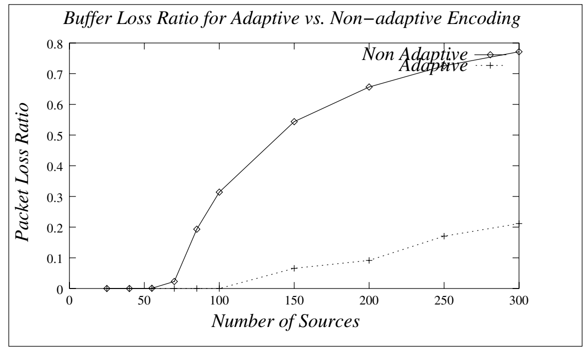

Fig. 10. Buffer Loss Ratio for Adaptive and Non-adaptive Encoding

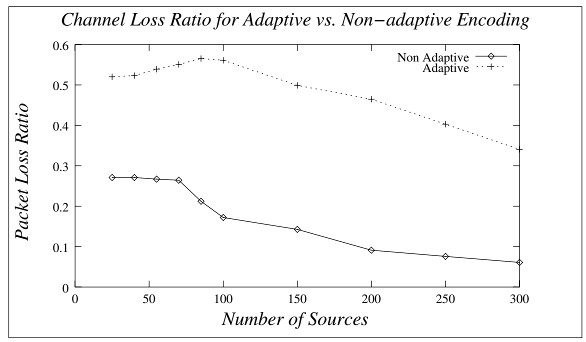

Fig. 11. Channel Loss Ratio for Adaptive vs. Non-adaptive Modulation.

buffer losses), some of these packets will have to face loss in the channel, and that is what we see in Figure 8. Now looking back at Figure 6, we can understand its behavior. For number of sources less than 85, both schemes had 0 buffer losses, but the adaptive scheme had higher channel losses, and that's the reason for higher overall loss ratio in figer 6. As the number of sources starts to increase, the adaptive approach will have less overall loss ratio. Even though adaptive modulation achieved more multiplexing, and better overall loss performance, still a high loss ratio occurs in the channel.

However, using any error correction mechanism will allow us to go around this kind of loss. Figure 9 shows adaptive versus non-adaptive encoding, where in both cases, adaptive modulation is used (case 2 above). An interesting thing that we notice in the graph is the stability of losses of the adaptive approach, where the loss ratio is around 50% regardless of the number of sources, as compared to the non-adaptive approach where the loss ratio increases as the number of sources increases. This indicates that the scheme succeeds in finding a suitable operating point (trading off multiplexing gains and voice quality). The reason for this stability will be explained in the coming two graphs.

Figure 10 shows the packet loss ratio due to buffer tail drops. The loss ratio for the adaptive approach is constantly much less than that for the non-adaptive one. This reflects that the adaptive scheme allows for multiplexing gain, and the ability to support more voice communications. As we mentioned, we are planning to use an error correction mechanism to go around the excessive number of channel losses. Figure 11 shows the packet loss ratio due to the wireless channel errors. Since buffer losses for the adaptive approach were much less than those for the non-adaptive, we expect the channel losses to be higher, and that is what Figure 11 shows. In the case of the adaptive scheme, the increase in the buffer losses was faced in a mostly equal decrease in the channel losses, and this is the reason why the graph in Figure 9 is stable. The non-adaptive aproah, on the other hand, has a very high buffer loss ratio that affects the overall loss ratio. Now that we are under the impression that adaptive voice provides better performance than the non-adaptive one, we wanted to quantify the quality achieved in terms of packet loss ratio and delays, which resemble the main metrics for voice quality. A voice packet that arrives "too" late to the destination will be dropped as it cannot be played out of order. In Figure 12, we consider any packet that waits for more than 80 ms in the buffer to be stale. Again, the adaptive approach has a superior behavior.

Figure 13 shows the degradation in voice quality (DVQ) for the adaptive versus the non-adaptive approach. DVQ is defined as:

$$\text{DVQ} = \frac{\text{Nos. packets lost} + \text{Nos. packets above delay threshold}}{\text{Total Number of packets}}.$$

As can be seen from figure, the adaptive approach has a much less DVQ. Hence it provides better quality and higher number of voice communications. In the non-adaptive approach, how-

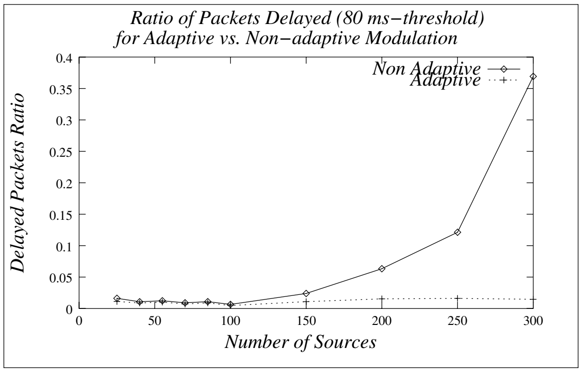

Fig. 12. Ratio of Packets Delayed beyond 80 ms threshold for adaptive and non-adaptive modulation

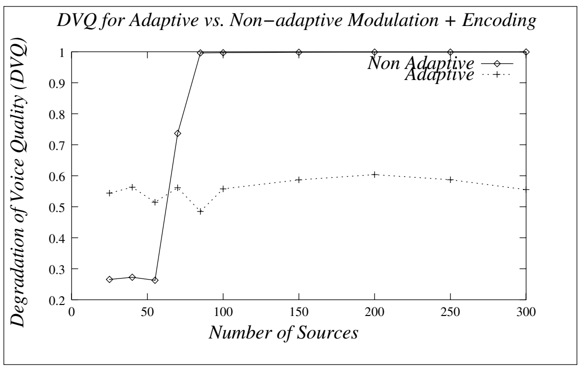

Fig. 13. Degradation in Voice Quality for Adaptive vs. Non-adaptive Modulation

ever, after reaching 85 sources, most of the load will be unuseful due to the fact that packets are either lost due to error channels or delayed long times in the buffer.

G. Conclusions

We studied the performance of adaptive voice over one-hop communication. Adaptability was achieved at two levels: (i) changing the voice-encoding rate, and (ii) changing the signal modulation scheme. Results show that the adaptive approach has a less degradation in voice quality (DVQ) and allows for higher number of voice sources to be supported. Currently, we are investigating generalizing this approach by having a multiple state wireless channel, and multiple modulation and encoding rates. We are also trying to use the multiplexing gain for error control. Once this is achieved, we will generalize the idea to a multi-hop context, where an ad hoc-aware MAC protocol is used. We also plan to evaluate the performance of the schemes using perceptual metrics such as ITU PESQ (perceptual speech quality) P.862.

Recent studies have shown that correlation between objective measures and subjective MOS (mean opinion score) is not very strong [18]. Hence our ultimate goal is to compare the adaptive and non-adaptive scheme using subjective measures.

IV. Acknowledgements

The authors would like to thank the anonymous refrees for their helpful suggestions which helped us in improving the quality of the paper. This research is supported in part by National Science Foundation Grants ANI-0086020 and ANI-0196156.

References

[1] C. PartRidge "Isochronous applications do not need jitter controlled networks". *RFC1257*, September, 1991.
[2] D. E. McDysan, and D. Spohn. "ATM Theory and Applications". *McGraw-Hill, Signature Edition*, 1999.
[3] R. Jain, S. Munir, J. Iyer. "Performance of VBR Voice over ATM: Effect of scheduling and Drop Policies". *ATM Forum/97-0608*, 1997.
[4] G. Holland, N. Vaidya, P. Bahl. "A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks". *ACM/IEEE Int. Conf. On Mobile Computing and Networking*, 2001.
[5] T. Hindelang, M. Kaindl, J. Hagenauer, M. Schmautz, and W. Xu "Improved Channel Coding and Estimation for Adaptive Multi Rate (AMR) Speech Transmission". *Vehicular Technology Conference Proceedings, 2000*, VTC 2000-Spring Tokyo. 2000 IEEE 51st , Volume: 2 , 2000
[6] S. Shenker "Fundamental Design Issues for the Future Internet". *IEEE Journal on Selected Areas in Communications*, Vol 13, NO. 7, pp. 1176-1188, 1995
[7] A. Barberis, C. Casetti, J. C. De Martin, M. Meo "A Simulation Study of Adaptive Voice Communications on IP Networks" *2000 Intl. Symposium on Performance Evaluation of Computer and Telecommunication Systems*, pp 531-542, July 2000
[8] K. Balachandran, S. R. Kadaba, and S. Nanda "Channel Quality Estimation and Rate Adaptation for Cellular Mobile Radio". *IEEE Journal on Selected Areas in Communications*, Vol. 17, NO. 7, pp. 1244-1256, 1999
[9] T. Ue, S. Sampei, N. Morinaga, and K. Hamaguchi "Symbol rate and modulation level-controlled adaptive modulation/TDMA/TDD system for high-bit-rate wireless data transmission". *IEEE Transactions on Vehicular Technology*, Vol 47, NO. 4, pp. 1134-1147, 1998.
[10] D. L. Goeckel "Adaptive coding for time-varying channels using outdated fading estimates Algorithms". *IEEE Transactions on Communications*, Vol. 47, NO. 6, pp.844-855, 1999
[11] Motorola Computer Group GSM-AMR Transcoder based on FACT.
[12] D. Minoli, and E. Minoli "Delivering Voice over IP Networks". *Wiley*, Wiley, 1998.
[13] P. T. Brady ""A Model for Generating ON-OFF Speech Patterns in Two-Way Conversations". *Bell System Technology Journal*, Vol. 48, pp 2445-2472, Sept. 1969.
[14] S. Deng "Traffic Characteristics of Packet Voice". *IEEE International Conference on Communications*, Vol. 3, pp 1369-1374, 1995.
[15] H. S. Wang and N. Moayeri "Finite-State Markov channel - a useful model for radio communication channels". *IEEE Transactions on Vehicular Tech*, Vol. 44, no. 1, Feb. 1995.
[16] E. N. Gilbert "Capacity of a burst-noise chanel". *Bell Systems Technical Journal*, Vol. 39, Sept 1960, pp. 1253-1266.
[17] E. O. Elliot "Estimates of error rates for codes on burst-error channels". *Bells Systems Technical Journal*, Vol. 42, Sept. 1963, pp. 1977-1997.
[18] L. Sun and E. Ifeachor "Subjective and Objective Speech Quality Evaluation under Bursty Loss". *Measurement of Speech and Audio Quality in Networks*, Online Workshop MESAQIN, 2002. (http://wireless.feld.cvut.cz/mesaqin/index.html).