Reliable Audio for Use over the Internet

Vicky Hardman <v.hardman@cs.ucl.ac.uk>, Martina Angela Sasse <a.sasse@cs.ucl.ac.uk>,
Mark Handley <m.handley@cs.ucl.ac.uk>, Anna Watson <a.watson@cs.ucl.ac.uk>

Abstract

This paper describes current problems found with audio applications over the MBONE (Multicast Backbone), and investigates possible solutions to the most common one - packet loss. The principles of packet speech systems are discussed, and how the structure allows the use of redundancy to design viable solutions to the problem. The paper proposes the use of synthetic speech coding algorithms (vocoders) to provide redundancy, since the algorithms produce a very low bit-rate stream, which only adds a small overhead to a packet. Preliminary experiments show that normal speech repaired with synthetic quality speech is intelligible, even at very high loss rates.

Introduction

The application of this work is multimedia conferencing over the MBONE (Multicast Backbone), an experimental overlay network of the Internet. The work has arisen from experiences in multi-way multimedia conferencing in Project MICE (Multimedia Integrated Conferencing for Europe) [1], is currently applied in Project ReLaTe (Remote Language Teaching over SuperJANET) [2], and includes formal experiments into the human perception of packet speech systems degraded by packet loss.

If multimedia conferencing is to become widely used in the Internet community, user must perceive the quality to be sufficiently good for most applications. Experience has shown that audio is almost always the most important component of multimedia conferencing. Whilst we have identified a number of problems which impair the quality of audio, the major one with audio over the MBONE is packet loss [3]. This paper attempts addresses the problem of packet loss over the MBONE.

Packet loss can occur for a number of reasons:

- congestion of routers and gateways, which lead to packet being discarded;
- delays in packet transmission, with packet arriving too late at the receiver to be played back;
- heavy loading of the workstations, leading to scheduling difficulties in multi-tasking operating systems.

Packet loss is a persistent problem, particularly given the increasing popularity, and therefore increasing lead, of the Internet. Possible ways of combatting congestion include bandwidth reservation and moves toward an integrated service management on the Internet. These would require wide-scale changed to be agreed and implemented, so these solutions will be available in the short to medium term. Yet, the disruption of speech intelligibility even at low loss rates which we currently experience may convince a whole generation of users that multimedia conferencing over the Internet is not viable. We therefore propose a solution which renders the speech intelligible under current network conditions, and can be deployed in the short term. Such a solution will have to be at the application level, i.e. the multicast audio tools.

Current audio applications repair lost packets with silence, which leads to the speech clipping effects currently experienced by many users. Since comparatively large packets are used, even the loss of individual packet loss has a serious impact on the intelligibility of speech.

We propose a method of repairing damaged speech using cheap redundancy within the packets sent from the transmitter. The redundancy is synthetic speech, which, when split into packets, only adds a very small amount of overhead, and therefore does not add to the congestion at the network level. The redundancy for any given packet of speech is piggybacked onto a later packet. This mechanism means that when the receiver suffers the loss of the primary speech information, it still has the possibility of substituting something sensible in the output stream of speech, provided that the redundancy can be received.

In order to establish the effectiveness of this solution, we have performed experiments into user perception of speech repaired with a synthetic substitute. The experiments subjectively measured speech intelligibility, and the results show that this technique is very successful at repairing speech with large packet sizes and for very high loss rates (results were taken up to 40%). The paper also describes how the proposed solution scales in the multicast environment.

Background

Speech Coding

Speech coding schemes have been standardised for use over telephone networks; a variety of speech coding algorithms exist for a single target quality of service (QoS), and at a very few discrete bit-rates: Pulse Code Modulation (PCM) operates at 64 kbps, Adaptive Differential Pulse Code Modulation (ADPCM) operates at 32 kbps, and Code Excited Linear Prediction (LD-CELP) operates at 16 kbps. The target QoS is 'toll' (or telephone) quality, and each algorithm available at this QoS produces a different bitrate: the improvement in bit-rate being obtained for increasing complexity in the coding algorithms. Another standard speech coding algorithm is Groupe Speciale Mobile (GSM), which was designed for use over cellular telephone networks. The target QoS is consequently slightly less than toll, but the algorithm is popular, since it operates at the same bit-rate as CELP, but is much less complex. A fuller discussion of toll quality speech coding algorithms can be found in [4].

A second class of coding algorithms exist, which operate at the 'communications' or synthetic QoS. These algorithms operate at very low bit-rates (approx. 4.8kbps and below), and produce very mechanical sounding speech. Perhaps the most important method of this class is Linear Predictive Coding (LPC), since the principle is also an integral part of both the CELP and GSM coders. A fuller description of which can be found in [5].

Packet Speech Systems

Packet speech systems usually employ the standard speech coding algorithms, and group the emerging stream of codewords into packets for transmission over the network. At the receiver, the packets may be delivered: out of order, not at all, or at non-uniform intervals. Consequently, a reconstruction delay must be used at the receiver to repair the network effects; this enables sample play-out to be smoothed.

In a packet speech system, the end-to-end delay is always a critical factor in the usability of a real-time voice system, and should be kept below 600ms in the absence of echoes (The figure may be in fact be less than this - 400ms) [6], if conversation patterns are not to break down. The size of the packets (in ms) chosen for a packet speech system directly impacts the end-to-end delay. A delay equal to the size of one packet is incurred at the transmitter, since the samples in the packet have to be collected before a packet can be sent. At the receiver, a rough estimate of the reconstruction delay required to smooth out packet arrival times is two packets worth in ms [7] [8], although the true value may be substantially in excess of this rule of thumb. Consequently, a minimum of three packets worth of delay is incurred on an end-to-end basis, before the network propagation delay has been taken into account.

The delay introduced will be enough to receive most of the packets, but some will always arrive too late to be played back, and can be considered 'lost'. Furthermore, the network itself may lose packets. In such situations, the speech 'stream' must be repaired, and a dummy packet inserted in place of the lost one, so that the correct timing relationship is maintained between the transmitter and receiver. The presence of the dummy packet is usually discernible to the listener, and unfortunately, the perceptibility of the loss increases with increasing packet size, as well as with increasing loss rate.

The impact of the two factors identified above, (delay and loss), is such that small packet sizes are required for real-time voice links. However, the use of small packets increases the overhead of packet headers, and any processing incurred at network nodes, and therefore increases the likelihood of congestion and loss. Consequently, a trade-off exists between the requirements of the network, and the requirements of real-time voice connections.

Voice Reconstruction Techniques

Repair methods for packet loss are known as voice reconstruction mechanisms. The aim is to construct a suitable dummy packet at the receiver, so that the loss is as imperceptible as possible. With compressed speech, voice reconstruction mechanisms not only have to produce a suitable fill-in packet, but also have to maintain the decoder tracking, since the algorithms transmit difference information. Voice reconstruction techniques can be split into two categories; receiver only, and combined source and channel techniques.

Table 1: Voice Reconstruction Techniques

Receiver-Only	Combined Source and Channel
Silence	Embedded Coding
White Noise	Redundancy
Waveform Substitution	
Sample Interpolation	

Receiver-only techniques are those that try to reconstruct the missing segment of speech solely at the receiver, possibly from correctly received packets preceding that which was lost. Combined source and channel techniques are those that try to make the system robust to loss by either arranging for the transmitter to code the speech in such a way as to be robust to packet loss, or by transmitting extra information to help with reconstruction.

Receiver-OnlyTechniques

The original voice reconstruction techniques were receiver-only, and used either silence, white noise, repetition of part of the last correctly received speech waveform, or sample interpolation as the substitute.

Silence substitution is favoured because it is simple to implement, and it gives adequate performance for small packet sizes (<16ms), and up to 1% loss [9] [10].

It is well known that other methods give substantially better results than those obtained from silence substitution. Warren [11] investigated the human perception of speech interrupted by silence compared to noises, such as coughs. The results show that phonemic restoration (the ability of the human brain to subconsciously repair the missing segment of speech with the correct sound) occurs for the noise situation, and does not occur for silence substitution.

Experience from the MICE project has shown that listeners can become very frustrated with MBONE speech. Their frustration stems from a variety of audio problems:

- interrupted speech due to packet loss - the speech sounds at first bubbly, and then individual interruptions are apparent;
- talking into a 'dead' channel, and the lack of automatic gain control for listeners;
- high levels of background noise cutting in and out;
- distortion due to overloading the microphone when speakers shout, or mis-matched levels.

Packet loss is the most frustrating problem, and one users cannot cure of their own accord. The frustration with interrupted speech can be explained by considering the linguistic construct of a sentence, which includes a pause (of duration > a phoneme) at the end of the sentence. Since the size of packets used over the MBONE are often comparable to the length of a phoneme, the interruptions in the speech flow sometimes occur at inappropriate points, which sends ambiguous signals to brain, as to whether speech is continuing or not [11].

White noise was shown to give a subjective performance improvement over silence by Miller [12] when contextual information in speech was removed, and an intelligibility improvement by Warren [11] when the contextual information was present. Consequently, silence substitution is not a suitable means of voice reconstruction, since white noise is known to give improvements, and is as easy to generate as silence.

Other receiver-only voice reconstruction techniques rely on the assumption that the speech characteristics have not changed from a preceding segment of speech, and use this preceding segment information to reconstruct the missing part; a simple example of this sort of voice reconstruction would be to repeat the last correctly received packet.The mechanisms fail when the packet sizes are large, and the loss rate is high (packets are more likely to be lost in twos or threes, than singularly). A fuller explanation of existing receiver-only techniques can be found in [13].

Combined Source and Channel Techniques

Combined source and channel techniques generally show significant improvement over receiver only techniques. The techniques either transmit extra information within the speech packets (to help with reconstruction at the receiver), or alter the speech coding algorithm and network operation (to make system as a whole more robust to packet loss).

Embedded speech coding techniques used with adaptive differential pulse code modulation (ADPCM [14]), such as those by [15], [16], and code excited linear prediction (CELP) [17], have shown significant performance improvements during packet loss. Embedded speech coding techniques allow the bit-rate can be adjusted from 40 to 32 or 23 kbps, without the introduction of large amounts of noise; essentially the feed-back loops in the encoder and decoder operate at a lower resolution than usual. The standard was designed to ease the problem of packet loss in packet networks; the codewords are segmented into high and low priority bits, and then placed in different packets. The mechanism relies on arranging for the network to drop packets containing LSBs only, which means that the mechanism is not applicable to networks which do not provide this support, such as today's Internet.

Lara-Barron [15] investigated embedded ADPCM coding techniques at 16-32 ms, and reported success for up to 40% loss (no reduction in speech quality for up to 6% loss).

The significant improvement resulting from the use of this mechanism is mostly due to the preservation of the decoder adaption logic [16].

Speech Quality

Speech quality may be assessed by either subjective or objective means, although it is well known that subjective assessment methods provide more accurate results.

Subjective assessment is usually made by performing listening tests using a large number of subjects. The material used, and the measurements made, depend upon the likely degree of distortion expected.

Toll quality speech coding algorithms are usually assessed by mean opinion scores (MOS) [18], where encoding distortion and noise are the likely type of degradation suffered. The technique involves the listener making a category rating after listening to a passage of speech.

Synthetic quality speech coding algorithms result in speech that has far greater degradation than found in toll quality systems; intelligibility is usually only adequate at best. Consequently, the MOS method is not suitable, and communications quality systems are assessed using comprehension or intelligibility tests. There is a wide range of speech material available, ranging from a sequence of syllables (the listeners transcribe what they hear) to passages (the comprehension of which is ascertained by asking a series of questions) [19].

The speech material is chosen based on the required sensitivity of the results, desired experimental control, and range of human faculties included in the test.

Reliable Audio for Use over the Internet

We have developed a new voice reconstruction scheme that uses redundancy to improve voice reconstruction at the receiver.

The redundant information is the output of a synthetic quality speech coding algorithm (LPC), which is very low bit-rate (4.8kbps). LPC is generally considered to contain about 60% of the information content of the speech signal, as the overall shape of the frequency spectrum is preserved at the expense of short-term amplitude and pitch variations. This technique is exactly what is required for successful voice reconstruction; the gap will be filled with a sound that is expected, and phonemic restoration should improve the situation further.

The Characteristics of the MBONE / Internet

The Internet, and its multicast overlay (MBONE) is a unique 'shared' packet network, that offers scalable multi-way communication. Such a network has traditionally not been considered suitable for speech applications, because of the large end-to-end delay that is commonly experienced over the network, and the potentially high probability of packet loss, (with relatively large segments of speech being lost). Current audio tools such as vat [20], and nevot [21] have, however, demonstrated that these problems are not prohibitive to successful voice communications, since use of these tools is very widespread.

The Internet provides variable length packets, a feature which has the potential for fine-grained control over the trade-off between network and speech performance requirements. The 'per-packet', rather than 'size-of-packet' network penalty for small packets coupled with the ability to have variable length packets also means that the state information from the speech coding algorithms can be transmitted in the packet, which substantially improves the perception of packet loss. Current audio tools transmit coding algorithm state information in each packet, but replace lost packets with silence.

The Loss Characteristics of the MBONE

Current research by a MICE partner, Bolot [22], is investigating the number of consecutive losses found over the MBONE. The results show that for light and intermediate loads, losses are essentially non-consecutive for an audio stream, and for heavy loads, the behaviour is similar, but consecutive losses are more prevalent.

These results suggest that a model where the redundancy is positioned in the packet after speech from the primary coding algorithm is suitable for light and medium network loads, and a model with the redundancy positioned a number of packets later is suitable for heavy loads.

Voice Reconstruction for the MBONE

LPC as the redundant information adds only a small amount of overhead to an RTP [23] packet (12 bytes per 160 bytes of PCM (or per 80 bytes of ADPCM). The information is piggy-backed to the packet following that containing the primary speech codewords; that the loss of an individual packet can be repaired using the redundant information in the following packet. This mechanism is unique to packet networks, and is only feasible because of the reconstruction delay introduced at the receiver.

The use of this redundancy technique means an increase in the reconstruction delay by the time equivalent of the distance of the redundancy component after the primary component; this implies an extra delay of one packet for light and medium loading conditions.

Multiple multicast receivers in a single conference may experience a variety of the characteristics reported in [22]. Consequently, the reconstruction mechanism may occasionally have more than one instance of the redundancy after the primary coding scheme packet. In this way, the heavy loading characteristics seen by one site do not affect the performance of the majority.

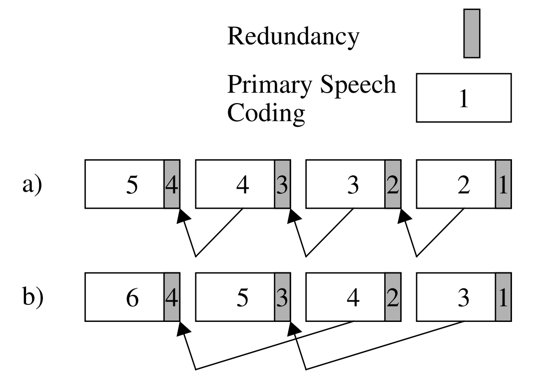

Figure 1: The Positioning of Redundancy Relevant to the Primary Coding for a) low and b) higher loss rates

The provision of LPC redundant information for use in voice reconstruction is intended to be used with per-packet state information; this prevents decoder mistracking in the case of loss. When a packet has been lost, the receiver decodes the redundant information, and feeds the samples to the audio hardware. Consequently, the output speech waveform consists of periods of toll quality speech, interspersed with periods of synthetic quality speech.

While LPC is a fairly complex speech coding algorithm, it should be noted that linear predictive analysis and synthesis are an essential part of all new higher compression schemes: GSM and CELP both use these techniques as a first step in their algorithms. LPC also has the potential to be used elsewhere in the system; as an improvement to the silence detection function, which is an integral part of most packet speech systems.

Experimental Design

Voice reconstruction experiments to date have usually been performed with packet sizes of 16-32ms, or less. The packet sizes used over the MBONE are usually greater than these values (40ms is commonly used). Consequently, little information exists about the degradation commonly experienced in voice connections over the MBONE/Internet. This experiment was therefore designed to compare LPC redundancy with waveform substitution and silence substitution; receiver-only techniques, such as waveform substitution are cheap to implement, and could potentially be used instead of LPC under certain circumstances. The experiment was also designed to try and give a broad outlook on the question of the degradation experienced as a result of packet loss over the Internet, by considering a wide range of loss rates (0-40%).

Waveform substitution was chosen as a representative receiver-only method, and the simplest of these, packet repetition, was considered suitable for comparison.

It was assumed that LPC redundancy could always be received in the event of loss - this assumption does not hold true in the real world, but provides a valid basis for the experiment design.

Waveform substitution repeats the last correctly received packet until the period of loss ends. This mechanism has an inherent draw-back when the lost packet is the last in a talk-spurt; the last packet will be repeated until a new talk-spurt starts.

The solution to this problem lies in realising that there should be a limit on the length of 'waveform substituted' speech, before the underlying assumption of no change in the speech characteristics breaks down; a suitable figure for this is 80ms, or the average length of a phoneme. Consequently, when the packet size is 20ms, the last correctly received packet may be repeated three times. When the packet size is 40 ms, the packet may be repeated twice (120ms). When the packet size is 80ms, the packet may be repeated once (160ms).

The speech was recorded and manipulated using a Sun SPARC station 10, the OGI speech tools software [24], which is a system that was developed for speech recognition experiments, and software written by the author (to generate the loss, code and decode the speech using different algorithms etc.). The codecs used in these experiments are publicly available versions [ADPCM][LPC], with ADPCM conforming to the CCITT standard [25]. The loss was generated randomly.

The subjective quality was assessed using phonetically balanced (PB) words, which proportionally represent the sounds found in every-day English speech.

There were three groups of seven subjects, those who heard the PB word lists reconstructed by silence substitution, those who heard waveform substitution, and those who heard LPC reconstruction. Each subject in each group had ten different lists, nine of which were test conditions, and the first of which was a no loss control condition. The lists were 25 words long, and after recording answers, for each list, the subjects completed a MOS rating scale, rating the quality of the speech just heard on a five-point scale, from bad to excellent.

A detailed description of the experimental details can be found in [13] and [26].

Results

The interaction between reconstruction scheme and packet size is difficult to analyse, since other factors, such as the temporal characteristics of speech sounds, and the masking and temporal order perception capabilities of the ear also affect human perception.

Silence Substitution

The first point to consider is whether the results from these experiments give an insight into the speech perception from current audio tools.

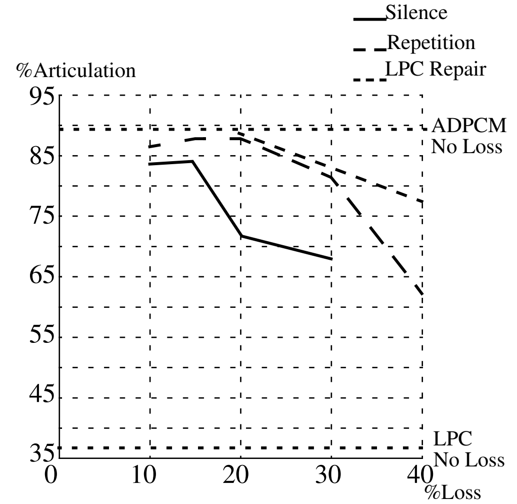

Figure 2: Voice Reconstruction for 20/40ms Packets

Referring to figure 2: Silence Substitution for 20 and 40 ms packet sizes, it can be seen that the graph indicates that using silence substitution for the voice reconstruction scheme fails between 15 and 20% loss. For 80ms packets (see figure 3), the results suggest that intelligibility is inadequate at a 15% loss rate.

This observation is consistent with experiences obtained from Project MICE. The observation is also in keeping with findings reported by [12] and [27]; 'speech degradation as high as 50% can be tolerated (intelligibility of 80%) if the packet size is small (0.019s)' on the other hand, if packets are long (0.25s), and the loss probability is high, intelligibility decreases to very low values (10%)'.

Waveform Substitution

The results for waveform substitution indicate that, for a packet size of 20ms, intelligibility does not decrease significantly with loss rate over the range of measurements taken. When the packet size is 40 ms, however, a significant difference is found between 30% loss and 40% loss, whilst with 80 ms packets significant differences can be found at much lower levels of loss: performance drops between 15 and 20%, and between 20 and 30%.

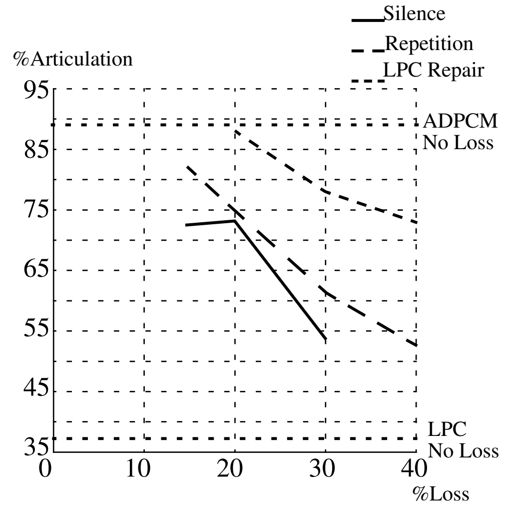

Figure 3: Voice Reconstruction for 80ms Packets

Comparison of Silence Substitution and Waveform Substitution

A comparison between the two receiver-only voice reconstruction schemes was made for different packet sizes. The results show that waveform substitution is better than silence substitution for packet sizes of 20 and 40ms, but that the advantage is not present for packet sizes of 80ms. The highest advantage was obtained for packet sizes of 20ms.

The reduction in advantage of voice reconstruction scheme for 40ms (compared to 20ms) might be due to the practice within the voice reconstruction scheme of allowing a 40ms packet to be repeated up to twice, which implies an assumption that speech characteristics have not changed for 80ms. This assumption is unlikely to be valid all of the time, since the average length of a phoneme is 80ms. Better results might be obtained by restricting repetition to one 40ms interval only.

LPC Redundancy

Referring to Figure 2, it can be seen that for packet sizes of 20 and 40ms, as loss rate increases, intelligibility does not deteriorate much. Slight deterioration in intelligibility is only present for packet sizes of 80ms, and then only at high loss rates.

Comparison of LPC and Waveform Substitution

For 20ms and 40ms packets, the results show that there is little advantage of using LPC instead of waveform substitution.

For 80ms packets, LPC should always be used.

MOS Results

The MOS results show the same trends as the intelligibility results. However, the listeners preference for LPC redundancy instead of waveform substitution for 40ms packets was more marked than is shown in the intelligibility graph (Figure 2).

Conclusions and Further Work

This paper has described a voice reconstruction method for use in packet networks. The mechanism is suitable for both unicast and multicast connections under all types of network conditions, although the main intelligibility advantage is to be found with large packet sizes, and medium to high loss rates; a receiver only voice reconstruction mechanism is suitable for light loss rates and small packet sizes. Listeners however, prefer LPC reconstruction for packet sizes of 40 and 80ms.

The mechanism may have minimal overhead in processor utilisation, since LPC analysis is part of the more complex coding algorithms, and may be used in the future to enhance other aspects of an audio tool. The method also only adds a small amount of overhead to the bandwidth used in a voice conference.

Since the overhead in terms of bandwidth is small, it may be desirable to have multiple copies of the redundancy present. This rationale would enable an audio tool to provide multi-cast international conferences with a voice reconstruction mechanism that scales; receivers at the end of 'good' network branches would have good performance with imperceptible loss, while receivers at the end of 'poor' network branches would experience a small reduction in speech quality, but would have a larger end-to-end delay.

The paper describes formal speech quality tests on the human perception of speech at different packet loss rates. The results show that the perceptibility of packet loss need no longer be regarded as one of the main constraints on the packet size if LPC redundancy is used to aid voice reconstruction.

In particular, the voice reconstruction should take the following form:

At low loss rates (20% and below) and small packet sizes (20 and 40ms) the results show that a receiver only technique is suitable, although LPC redundancy should also be used for 40ms packets, since listeners preferred it to waveform substitution.

At higher loss rates and for all conditions when packet size is 80ms, LPC redundancy should be used.

The MICE project is currently implementing a prototype audio tool, which will include voice reconstruction using LPC redundancy. Techniques to provide controlled feed-back from multiple receivers in multi-cast conferences are also being developed, which will enable the use of redundancy to be adaptive to receiver's needs. Further studies on human perception of audio and network performance, are planned, which will investigate the limits on packet size; the study will include an analysis of the impact of delay, as well as the human perception of large packet loss. The human perception studies will also address the relationship between objective measures (such as network performance results), and subjective perception and performance.

Acknowledgements

The ideas relating to using LPC for redundancy have been discussed at length in the technical meetings of the MICE (Multimedia Integrated Conferencing for Europe - ESPRIT 7606 Project). In particular, we would like to acknowledge the collaboration with Christian Huitema and Jean Bolot from INRIA Sophia Antipolis, France, on this issue. Very special thanks are due to Van Jacobson (Lawrence Berkley Labs), the creator of vat, for many fruitful discussions and email exchanges on audio problems and how to cure them.

References

[1] Kirstein P.T., Sasse M.A., Handley M.J.,'Recent Activities in the MICE Conferencing Project' Paper No. 166, INET 95.

[2] Buckett J. Campbell I. Watson T.J. Sasse M.A. Hardman V.J. Watson A. 'ReLaTe: Remote Language Teaching over SuperJANET' Proceedings of UKERNA 95, Networkshop, March 1995.

[3] Sasse M. A. et al. 'Remote Seminars Through Multimedia Conferencing: Experiences from the MICE Project, Proceedings of INET94/JENC5, pp. 251-258.

[4] Papamichalis P.E. 'Practical Approaches to speech Coding' Publ. Prentice-Hall 1987.

[5] Gold B. 'Digital Speech Networks' Proceedings of the IEEE, Vol. 65, No. 12, December 1977.

[6] Brady P.T. 'Effects of Transmission delay on Conversational Behaviour on Echo-Free Telephone Circuits' Bell System Technical Journal, pp 115-134, January 1971.

[7] Ades S.A. 'An Architecture for Integrated Services on the Local Area Network' Ph.D Thesis, Cambridge University Technical report 114.

[8] Jacobson V. 'Multimedia Conferencing on the Internet' Tutorial 4, presented at SIGCOMM 94, London, August 1994.

[9] Jayant N.S. & Christensen S.W. 'Effects of Packet Losses in Waveform Coded Speech and Improvements Due to Odd-Even Sample-Interpolation Procedure' IEEE Transactions on Communications, Vol. COM-29, No. 2, February 1981.

[10] Gruber J.G. & Strawczynski L. 'Subjective Effects of Variable Delay and Speech Clipping in Dynamically Managed Voice Systems' IEEE Transactions on Communications, Vol. COM-33, No. 8, August 1985.

[11] Warren R.M. 'Auditory Perception' Pergamon Press Inc.

[12] Miller G.A. & Licklider J.C.R 'The Intelligibility of Interrupted Speech' Journal of the Acoustical Society of America 22(20:167-173 (1950).

[13] Hardman V.J., Sasse M.A., Watson A. 'Successful Voice Reconstruction for Packet Networks using Redundancy' Research Note, Dept. of computer Science, University College London April 1995. RN/95/.

[14] CCITT G.727 '5-, 4-, 3- and 2-Bits Sample Embedded Adaptive Differential Pulse Code Modulation (ADPCM)' CCITT Fascicle III.4 - Rec. G.727.

[15] Lara-Barron M.M. & Lockhart G.B. 'Speech Encoding and Reconstruction for Packet-Based Networks' IEE Colloqium on Coding for Packet Video and Speech Transmission, Vol. 199, (3) pp. 1-4, 1992.

[16] Kitawaki N. & Nagabuchi H. 'Evaluation of Coded Speech Quality Degraded by Cell Loss in ATM Networks' Electronics and Communications in Japan, Part 3, Vol. 75, No. 9, 1992.

[17] Yong M. 'Study of Voice Packet Reconstruction Methods Applied to CELP Speech Coding' IEEE Journal on Acoustics, Speech and Signal processing, CAT No.92CH3103-9 Vol. 2, pp 125-128 1992.

[18] CCITT 'Recommendations of the P Series', 'Method for the Evaluation of Service from the Standpoint of Speech Transmission Quality' CCITT Red Book Volume V - VIIIth Plenary Assembly, 1984.

[19] Kryter K.D. 'Speech communication' Chapter 5 from Human Engineering guide to Equipment Design, Editors Van, Cott and Kinkade

[20] Jacobson V. 'VAT manual pages', Lawrence Berkeley Laboratory (LBL) February 1992.

[21] Schulzrinne H. 'Voice Communication Across the Internet: A Network Voice Terminal' University of Massachusetts, Technical Report, June 1992.

[22] Bolot J. Crepin H, Vega Garcia A. 'Analysis of Audio Packet Loss on the Internet', Proceedings NOSSADV 95 (Network and Operating System Support for Digital Audio and Video), pp 163-174, Durham, NH, April 95.

[23] 'RTP: A Transport Protocol for Real-Time Applications' Work-in-progress Internet Draft, Audio-Video Transport WG, version 7, March 21 1995.

[24] CSLU. 'OGI Speech Tools User's Manual' Technical Report, Center for Spoken language Understanding, Oregon Graduate Institute, 1993.

[25] '32 kbit/s Adaptive differential Pulse Code Modulation (ADPCM)' CCITT Fascicle III.4 - Rec. G.721.

[26] Watson A. 'Loss of Audio Information in Multimedia Videoconferencing - An Investigation into Methods of Assessing Different Means of Compensating for this Loss' MSc Thesis, Dept. of Ergonomics, Faculty of Science, University of

London.

[27] Minoli D. 'Optimal Packet Length for Packet Voice Communication' IEEE Transactions on Communications Vol. COM-27, No. 3, March 1979.

Author Information

Vicky Hardman is a research fellow working on project ReLaTe at UCL, specialising in audio and speech over packet networks. She has also worked on a number of national and international research projects, such as project MICE and project Unison. She has a PhD in speech over packet networks from the Electronic and Electrical Engineering Department, Loughborough University of Technology.

Martina Angela Sasse has been a lecturer in Computer Science at UCL since 1990. She has worked on a number of national and international research projects on multimedia conferencing, and is currently the Project Manager of MICE and Principal Investigator of ReLaTe at UCL.

Mark Handley received his BSc in Computer Science with Electrical Engineering from UCL in 1988. Since 1991, he has been a Research Fellow, working on the RACE CAR multimedia conferencing project and on subsequently on the MICE project, of which he is the Technical Director at UCL. .

Anna Watson is a research assistant on project ReLaTe in the Department of Computer Science at UCL, working on the ReLaTe. She has a degree in psychology and an MSc in Ergonomics/HCI.