Packet Loss Concealment for G.722 using Side Information with Application to Voice over Wireless LANs

Niranjan Shetty and Jerry D. Gibson
Department of Electrical and Computer Engineering, University of California, Santa Barbara, USA
Email: {niranjan, gibson}@ece.ucsb.edu

***Abstract*— The G.722 wideband speech codec offers higher quality and better naturalness than G.711, is low in complexity, has low delay, and tandems well with other codecs. This makes it an attractive codec for Voice over IP and Voice over Wireless LANs. However, loss of a G.722 coded speech frame results in a mismatch of the encoder/decoder states that affects the decoding of subsequent correctly received frames. We present a new packet loss concealment (PLC) method for G.722 and propose an efficient approach to sending the side information to improve concealment of the lost frame and to resynchronize the encoder and decoder. The proposed method greatly improves the robustness of the G.722 codec to packet losses for Voice over Wireless LANs. Wideband Perceptual Evaluation of Speech Quality Mean Opinion Score (WB-PESQ MOS) is used as the quality metric for evaluating the performance of the PLC schemes.**

***Index Terms*— G.722, packet loss concealment, voice over wireless LANs**

I. Introduction

With the wider passband of 50-7000 Hz, wideband speech provides much better quality and naturalness compared to narrowband telephone bandwidth speech. The G.722 codec encodes wideband speech by first decomposing it into two subbands, and then using backward adaptive differential pulse code modulation (ADPCM) to code each subband. At the receiver, the coded speech in each subband is decoded using an ADPCM decoder, and reconstructed to give the decoded wideband speech [1]. This simple coding and decoding operation causes an algorithmic delay of only 3 ms and provides good performance at rates of 48, 56 and 64 kbps [2], [3]. The low delay, low complexity and high quality of the G.722 wideband speech codec have resulted in its adoption by several Voice over IP (VoIP) and Voice over Wireless LAN (VoWLAN) phones [4]–[6].

For VoIP and VoWLAN applications, it is important to maintain good conversational quality in the presence of packet losses. Packet losses in VoWLANs can be attributed to impairments such as delay, channel errors, and collisions, which can cause the affected packets to be discarded by the receiver. Speech codecs rely on frame loss concealment schemes to reduce the degradation in perceptual quality caused by packet losses.

PLC schemes can be classified as coder-independent or coder-dependent [7]. The pitch repetition scheme used in the ITU-T G.711 codec PLC [8] is an example of a coder independent scheme that can be applied to all waveform codecs. These coder-independent PLC schemes have the advantage of wider applicability, but may be limited in their ability to resolve distortions during packet losses that stem from the specific coding technique employed. In some CELP based codecs like the AMR [9], and in the case of ADPCM based codecs like G.722, the correct decoding of each frame relies on the successful reception of the previous frame. As such, the occurrence of a frame loss affects the decoding of subsequent correctly received frames. In such instances, some additional information about the lost frame can not only be used for better concealment of the lost frame, but also for proper decoding of the subsequent frames. The use of additional side information as a means for improving concealment performance has been demonstrated for CELP-based coders such as G.729 [10] and AMR-WB [11].

A simplistic approach to concealing a frame loss for G.722 coded speech is to use pitch period repetitions of a previous correctly received frame as the concealment frame, as in case of the G.711 ITU PLC [8]. However, in the G.722 codec, the loss of a frame also desynchronizes the ADPCM decoders, resulting in a state mismatch between the encoders and decoders. As a result, when the next correctly received frame is processed, the lack of correct state information results in audible clicks in the decoded speech. In addressing such a mismatch caused due to channel errors in TDMA-TDD Personal Communication Systems, Kubota et al [12] suggested an algorithm that mutes the erroneous ADPCM speech samples if it crosses a certain threshold, and detects and suppresses the click-noise in the decoded PCM speech. A scheme for updating the internal state parameters was proposed by Serizawa and Nozawa [13], which involves updating the internal states of the ADPCM decoders by processing the concealment frame at the receiver.

This paper is based on "Improving the Robustness of the G.722 Wideband Speech Codec to Packet Losses for Voice over WLANs," by N. Shetty and J. D. Gibson, which appeared in the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. © 2006 IEEE.

This work was supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc. and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc.

While this method reduces the effect of clicks, their forgetting factor control tends to dampen the gains of the decoded speech excessively following a loss, resulting in a degradation of quality.

We observe that the state information needed to start coding a subband speech frame in the G.722 ADPCM encoder is the same as that required to begin decoding the coded subband speech frame in the ADPCM decoder at the receiver. Thus, if the state information for each subband can be sent to the decoder, the distortion due to state mismatch described earlier can be eliminated, providing improved PLC performance. Additionally, the use of the exact pitch value of the lost frame in generating the concealment frame, as opposed to an estimate of the pitch period from past decoded speech, is seen to improve the PLC performance. These observations formed the motivation for our proposed scheme that uses quantized side information for improving the PLC performance of the G.722 codec [14].

The side information is quantized using scalar and vector quantizers and the coded side information is then appended at the end of each speech frame in an IEEE 802.11 packet and transmitted to the decoder over the wireless link. The coded side information increases the overall 802.11 packet size by less than 7% and results in a significant improvement in PLC performance. When the work in [14] was performed, no standardized schemes existed for packet loss concealment in G.722. Since that time, the International Telecommunication Union (ITU) conducted a candidate submission process and has approved two packet loss concealment algorithms [15]. One of the submissions is described in [16]. The details of these algorithms are yet to be published.

This present paper is an expansion and elaboration of the work presented in [14]. The quantization schemes used for encoding the side information are presented in detail. Further, the application of the side information based PLC scheme to Voice over Wireless LANs is discussed. The operation of the G.722 codec is outlined in the next section. In Sec. III, we describe the *Modified Decoder Update PLC*, which is our modification to the scheme in [13] and results in an improvement in performance but does not eliminate the clicks due to state mismatch between the encoder and decoder states. In Sec. IV, we consider the case where perfect state side information and pitch information are available at the decoder for G.722 PLC. A method for quantizing the state information at the encoder for transmission to the decoder is investigated in Sec. V with the objective of achieving a voice quality as close as possible to the case with perfect side information. The implemented technique is referred to as the *Side Information Based PLC*. Finally, in Sec. VI, we compare the performance of the *Side Information based PLC* with the *Modified Decoder Update PLC* for Voice over Wireless LANs. We observe that, despite the slight increase in packet size due to addition of the side information for the *Side Information based PLC*, we obtain a significant improvement in PLC performance compared to the *Modified Decoder Update PLC*.

The speech quality is evaluated using an objective quality measure called the Wideband Perceptual Evaluation of Speech Quality (WB-PESQ) [17]. WB-PESQ is an extension to wideband speech of the PESQ [18], which is commonly used to evaluate the quality of narrowband speech. The consistency between subjective MOS scores and objective WB-PESQ was evaluated for G.722 [19] [20], and it was concluded that the WB-PESQ scores are consistent with the subjective MOS scores and provide a good assessment of perceived quality when the same coding conditions are used.

II. The G.722 codec [1], [21]

In the G.722 encoder, the wideband speech is first passed through 24-coefficient Quadrature Mirror Filters (QMFs) which decompose it into a lower subband component with frequencies between 0-4000 Hz and a higher subband component with frequencies between 4000-8000 Hz.

In each band, the predicted value $s[n]$ is subtracted from its input signal $x[n]$ to give the error signal $e[n]$. $s[n]$ is obtained by adaptive prediction as follows

$$s[n] = \sum_{i=1}^{2} a_{i,n-1} r_t[n-i] + \sum_{i=1}^{6} b_{i,n-1} d_t[n-i] \quad (1)$$

where $a_{i,n-1}$ and $b_{i,n-1}$ are the adaptive pole and zero coefficients, respectively, $d_t[n]$ is the quantized difference signal obtained by quantizing the error signal $e[n]$, and $r_t[n] = s[n] + d_t[n]$ is the quantized reconstructed signal.

The pole coefficients are updated as follows:

$$\begin{aligned} a_{1,n} &= \alpha a_{1,n-1} + 3(1-\alpha) sgn(p_t[n]) sgn(p_t[n-1]) \\ a_{2,n} &= \beta a_{2,n-1} + (1-\beta)(sgn(p_t[n]) sgn(p_t[n-2]) \\ &\quad - f.sgn(p_t[n]) sgn(p_t[n-1])) \end{aligned} \quad (2)$$

where α and β are the forgetting factors for each pole coefficient and have default values of 255/256 and 127/128, respectively, and f is a function of $a_{1,n}$. $p_t[n]$ is the partially reconstructed signal and is obtained by adding the difference signal $d_t[n]$ with the zero section of the predicted signal $s[n]$, and $sgn(.)$ refers to the sign of the enclosed parameter. The zero coefficients are updated using the equation

$$b_{i,n} = (1-2^{-8}) b_{i,n-1} + 2^{-7} sgn(d_t[n]) sgn(d_t[n-1]) \quad (3)$$

for i=1,..,6.

Since error sensitivity increases with increasing pole prediction order, a second order pole predictor is used above [1], [21], [22]. Further, the stability of the pole coefficients is ensured by the following constraints

$$|a_{2,n}| \leq 0.75$$

$$|a_{1,n}| \leq (15/16) - a_{2,n}$$

The logarithm of the scale factors associated with the adaptive quantizer for each subband $\nabla[n]$ are updated using the equation

$$\nabla(n) = \gamma\nabla(n-1) + W[I(n-1)] \tag{4}$$

where γ is the leakage constant equal to 127/128, and W is the logarithmic scaling factor multiplier whose values depend on the encoded codeword $I(.)$.

The steps involved in ADPCM decoding are similar to those described above for the encoder. The decoded output for each subband is then reconstructed using the receive QMF filters and summed to give the decoded wideband speech.

III. Modified Decoder Update Scheme

In Serizawa and Nozawa [13], the lost frame is first concealed at the receiver by using pitch-period repetitions of the previous correctly received frame in a manner similar to the ITU PLC [8] scheme for G.711. In this scheme, if the current frame is lost, then the pitch period is estimated from prior decoded speech, and pitch period repetitions of previously decoded samples are then used to generate the concealment frame. The PLC scheme in [8] is adapted for wideband speech by altering the buffer lengths and pitch search range appropriately. We refer to this concealment method as the *Pitch Repetition* scheme. For updating the state information, the concealment frame in [13] is passed through the send QMF filters and the ADPCM encoders for each subband at the receiver, and after that the ADPCM decoder states are updated with the resulting internal states of the ADPCM encoders. Additionally, the forgetting factors associated with the adaptive quantizer and pole predictor coefficients in the ADPCM decoder are altered for decoding the first 5 ms of the next correctly received frame. This is done with the objective of reducing the impact of the past values on the current forgetting factors. However, the value chosen in [13] for the forgetting factor γ of the adaptive quantizer excessively damps the gain of the reconstructed signal and results in worse performance than the pitch repetition scheme alone for certain packet loss patterns. We refer to the scheme in [13] as the *Decoder Update* scheme. In our modified version of the scheme, called the *Modified Decoder Update*, the quantization scaling factor is left unaltered from G.722 while the forgetting factors in Eq. (2) for updating the pole coefficients are altered to the values specified in [13]. The values for the forgetting factors are shown in Table I.

TABLE I.
Forgetting factor values for the different schemes

Method	α	β	γ
Default	255/256	127/128	127/128
Decoder Update	254/256	253/256	123/128
Modified Decoder Update	254/256	253/256	127/128

Figure 1 shows a segment of voiced speech decoded using the pitch repetition, decoder update, and modified decoder update schemes, respectively. The highlighted portions indicate the frame that is lost and has been concealed using each of the fore-mentioned methods at the receiver. The circled portions indicate the correctly received frame following the lost frame. In the *Pitch Repetition scheme*, Fig. 1b), we observe a sudden spike in the frame following the lost frame, caused by the state mismatch between the encoder and decoder states, as discussed earlier. The Decoder update scheme in Fig. 1 c) is successful in removing the spike, but ends up damping the gains significantly. The *Modified Decoder Update* PLC in Fig. 1 d) removes the spike and also manages to leave the gain unaffected in the frame following the lost frame. However, in listening to several processed speech files decoded using each of the forementioned schemes, we noted that the *Modified Decoder Update* PLC reduces the effect of certain clicks due to state mismatch between

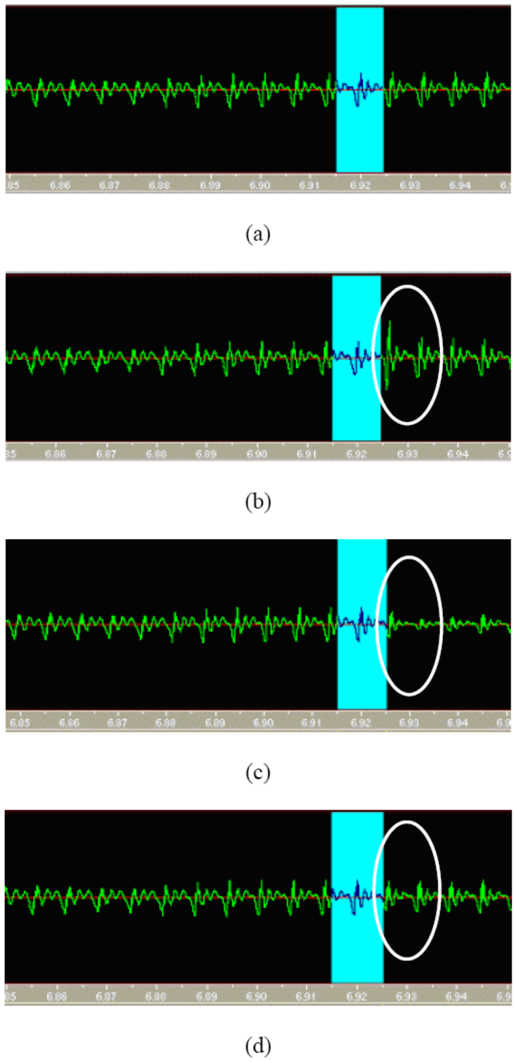


Figure 1. Waveforms for a) Reference decoded speech (no loss), b) Speech decoded using pitch repetition c) Speech decoded using the decoder update scheme d) Speech decoded using the Modified Decoder update scheme. The highlighted portion represents the lost frame, and the circled portion represents the correctly received frame following the lost frame

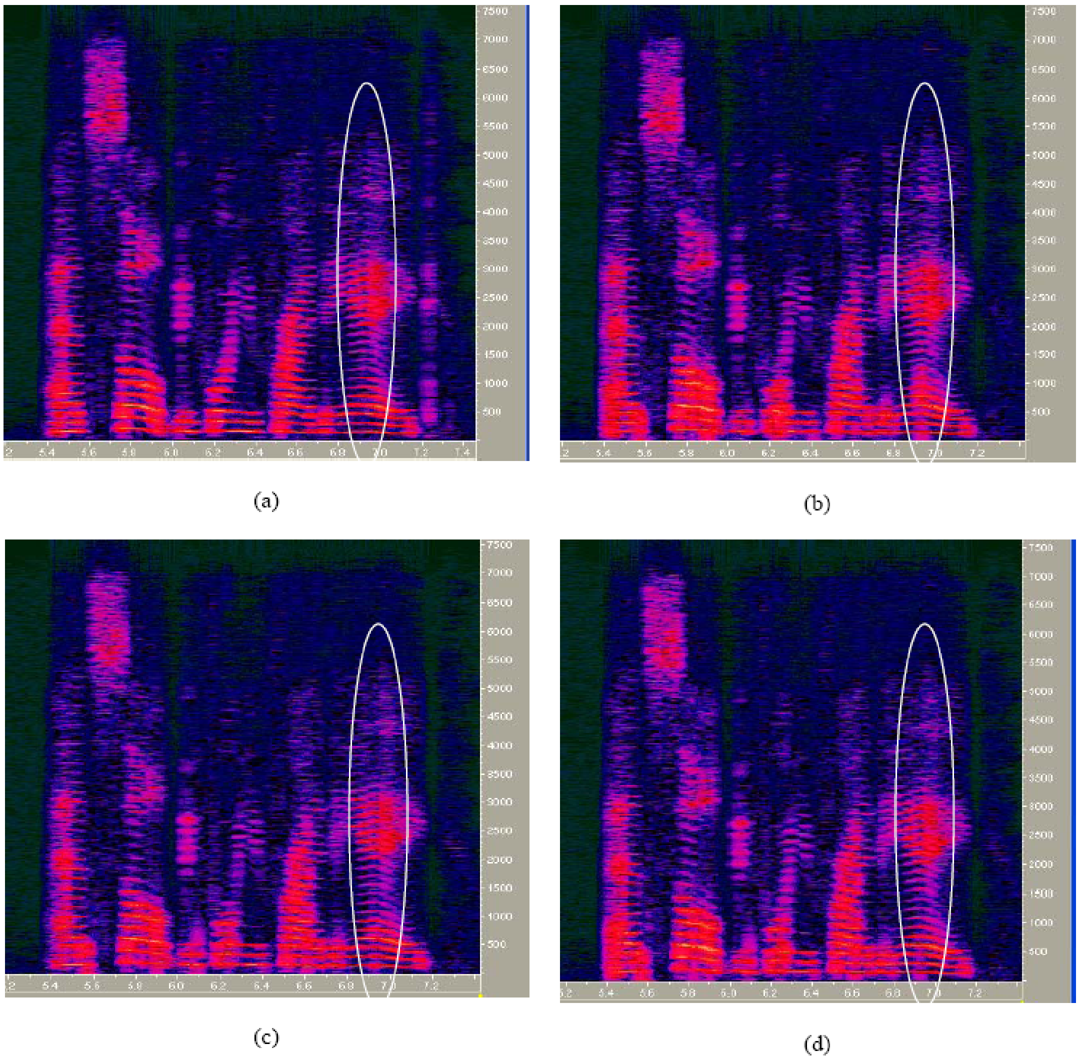


Figure 2. Spectrograms for a) Reference decoded speech (no loss) b) Speech decoded using pitch repetition c) Speech decoded using the decoder update scheme d) Speech decoded using the Modified Decoder update scheme. The circled region indicates the part of the spectrogram corresponding to the portion of the speech waveform shown in Fig. 1.

the encoder and decoder states but was still not rigorous enough to improve the performance significantly relative to the pitch repetition scheme.

The spectrograms for the forementioned cases are shown in Fig. 2. The circled region indicates the part of the spectrogram corresponding to the portion of the speech waveform shown in Fig. 1. In comparing the circled regions for each scheme, we observe a distinct low frequency distortion in the case of the Pitch Repetition scheme in Fig. 2 b) relative to the spectrogram of the decoded speech under no losses in Fig. 2 a). This distortion is observed between 6.9 and 6.95 seconds, and in the frequency band of 50-1500 Hz. This distortion is relatively deemphasized and restricted to a frequency band of 50-1000 Hz in the Decoder Update scheme in Fig. 2 c), and the pitch harmonics in this specified region of the spectrogram are much more well-organized than those in Fig. 2 b) for the Pitch Repetition scheme. In the Modified Decoder Update scheme in Fig. 2 d), a further improvement is observed, with the distortion restricted to a smaller frequency band of 50-700 Hz and with much more accurate reconstructed pitch harmonics. Despite these improvements, the spectrogram of the Modified Update Scheme in Fig. 2 d) still shows some distortions at low frequencies in the concealed regions when compared to the spectrogram in Fig. 2 a) for the no loss case. Our informal listening tests reveal that these distortions are indeed audible in the reconstructed speech. As a consequence, these modifications alone were not sufficient for acceptable PLC.

IV. PLC USING PERFECT SIDE INFORMATION

As a solution to updating the state information at the decoder, we consider another approach in which the state information from the encoder is available as side information at the decoder. We observe that for the ADPCM encoder in each subband, once the n^{th} frame is completely encoded, the state information needed to begin encoding the $(n+1)^{th}$ frame is available before that frame is encoded. This is also the same information that is needed to begin decoding the $(n+1)^{th}$ frame at the receiver. Thus, if the state information for each subband can be sent to the decoder, the clicks due to state mismatch described earlier may be eliminated, providing an improved PLC performance.

We first determined the best possible PLC performance with the availability of perfect state information as side information at the ADPCM decoders. The state side information required for each subband decoder is shown in Table II.

TABLE II.
STATE INFORMATION NEEDED AT THE ADPCM DECODER PER SUBBAND FOR RECONSTRUCTING SAMPLE AT TIME INSTANT n

$a_{i,n-1}$; $i = 1, 2$	Pole predictor coefficients
$b_{i,n-1}$; $i = 1, .., 6$	Zero predictor coefficients
$\nabla[n-1]$	Logarithmic quantizer scale factor
$p_t[n-i]$; $i = 1, 2$	Partially reconstructed signal
$r_t[n-i]$; $i = 1, 2$	Quantized reconstructed signal
$d_t[n-i]$; $i = 1, .., 6$	Quantized difference signal

The experiments were performed using two wideband speech files of male and female speech as input. Each segment is comprised of two sentences by the same speaker and is 8 seconds in duration. Each input file was pre-filtered based on the P.341 specification [23]. The default operating rate of 64 kbps is used for the G.722 codec. Each packet comprises a G.722 coded speech frame of duration 10 ms and size of 80 bytes. The packet loss rates (PLRs) considered were 0.1%, 0.5%, 1%, 5%, and 10%. For each PLR, 500 runs of the experiment were performed to simulate different packet loss patterns in the speech files. The average WB-PESQ calculated over the 500 different packet loss patterns corresponding to each PLR has been used as the quality measure in our experiments. The use of WB-PESQ is based on the results in [20] and [12] that have verified the suitability of using WB-PESQ for evaluating wideband speech quality coded using G.722 for packet loss rates up to 10%.

In Fig. 3 we compare the performance of the pitch repetition scheme based on [8], the modified decoder update scheme described in Sec. III, and the side information scheme which provides raw uncoded state information to each of the subband ADPCM decoders. Here, all of the methods use estimated pitch from the last decoded frame. We observe that the availability of perfect side information results in a significant improvement in the performance of the G.722 speech codec for a PLR as low as 1%, as compared to the pitch repetition scheme and the modified decoder update scheme.

As a next step, we investigated the contribution of the state information for each subband to the PLC performance. Following that, we investigated the effect on

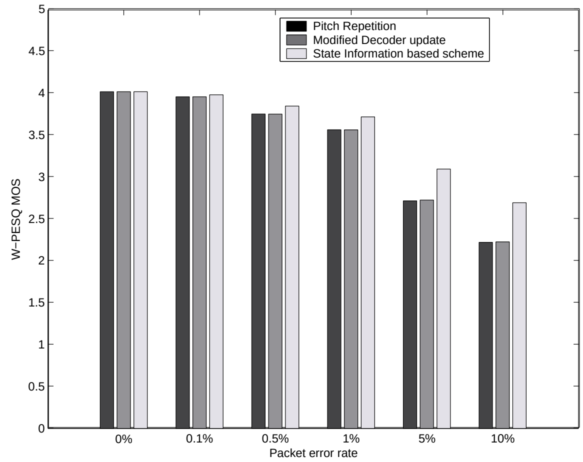


Figure 3. Performance of G.722 PLC using 1) Pitch repetition 2) Modified decoder update + pitch repetition 3) State side information + pitch repetition, all with estimated pitch

performance of using the correct pitch value instead of the estimated pitch value of the prior frame. These are discussed in the subsections that follow.

A. Effect of state side information for each subband on PLC performance

We investigate the effect on performance of providing state information for a single subband (lower or upper) relative to the performance when the state information for both the subbands is provided. In decoding the received frame following a lost frame, if the state information for a particular subband is available, the ADPCM decoder for that subband updates its parameters using the available state information. Pitch values estimated from the prior decoded frame are employed. The average WB-PESQ MOS values for the cases when the state information is available for the lower subband only, for the higher subband only, and for both the subbands, are compared with the pitch repetition scheme in Fig. 4.

We observe that communicating the state information for the lower subband alone results in an improvement in performance comparable to that obtained when state information from both the subbands is available at the decoder. Further, the availability of state information for the higher subband alone results in a performance comparable to the pitch repetition scheme. This implies that the state mismatch in the higher subband following a packet loss has negligible effect on the degradation in speech quality but that the state mismatch for the lower subband plays a significant role in the resulting degradation following a packet loss. This was further corroborated by informal listening tests.

Based on this result, we concluded that it is only necessary to provide the state information corresponding to the lower subband alone, thus reducing the amount of side information by 50%. In experiments that follow, the state information for the lower subband ADPCM alone is assumed to be available.

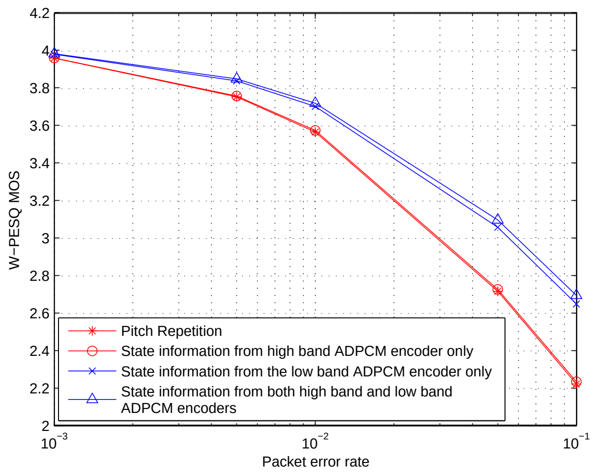


Figure 4. Performance comparison of G.722 PLC with various state information for individual subbands without pitch side information

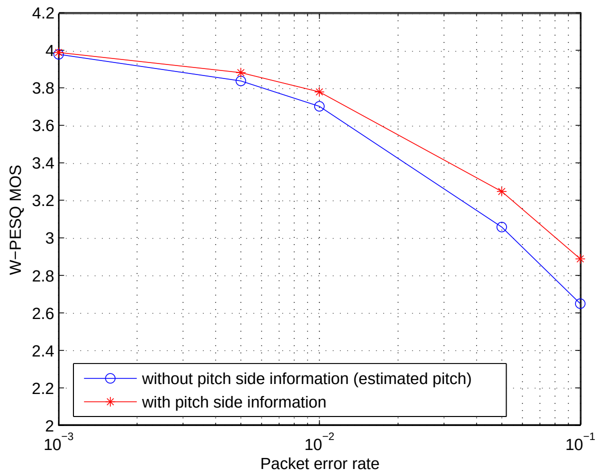


Figure 5. Performance comparison of Side Information based scheme with and without accurate pitch information.

B. Effect of pitch side information on PLC performance

The pitch repetition method used for generating the concealment frame estimates the pitch value from the previous correctly received frame, and uses pitch period repetitions of the previous frame to conceal the lost frame. Our next objective was to find out whether an improvement in performance can be gained if the true (not estimated) pitch value of the lost frame is available at the decoder, and is used instead of the estimated pitch value for repetition. For the case where the state side information is available at the decoder, we compare the performance of the G.722 PLC when the accurate pitch value is additionally available as side information to the performance when the pitch value is estimated from the previous correctly received frame as in the basic pitch repetition scheme. The PLC performance when only the state information is available at the decoder with the pitch values estimated from the previous frame is compared with the PLC when both state information and accurate pitch information for the lost frame were available at the decoder in Fig. 5.

From Fig. 5 we observe an improvement in performance when the accurate pitch for each frame is included in the side information. At a PLR of 10%, the availability of an accurate pitch value of the lost frame results in a MOS improvement in PESQ-MOS of 0.3. Thus, an additional improvement in performance can be obtained if the side information available at the decoder includes the accurate pitch values for the speech frames in addition to the state information.

V. Quantization of the Side Information

In the previous section we observe that the availability of accurate state and pitch information at the decoder results in an improvement in G.722 PLC performance. As the next step, we investigate a method for communicating the state and pitch information as side information to the decoder. This makes it necessary to quantize the side information at the encoder.

A method for quantizing the side information comprised of the state information specified in Table II and the pitch information, is investigated next. For the state information, the two pole coefficients in Eq. (2) were first converted into LSF coefficients to ensure a stable response after decoding at the receiver. The VQ design is based on the LBG algorithm with uniform divisive clustering [24]. For the VQ-codebook design, the training data was obtained by processing two male and female speech files through the ADPCM encoders and extracting the state information for each sample. The pole coefficients in the state information were converted into LSF coefficients and were used for training the VQ-codebook for the LSF coefficients. This resulting training data contained 128,000 LSF vectors. The training data and VQ codebook design for the zero coefficients and scaling factor were obtained similarly. The VQ-codebooks were trained using two different male and female speech files, each of duration 8 seconds, and then used for coding side information for another set of speech files from different male and female speakers than those used for training.

The mean square distortion in the LSF coefficients, and the zero coefficients and scaling factor corresponding to different quantization levels are shown in Fig. 6 and Fig. 7 respectively. In Fig. 6, corresponding to the VQ design for the LSF coefficients, we observe that the distortion curve has a very small slope for VQ levels of 6 bits = 64 codewords and higher. Similarly, in Fig. 7, corresponding to the VQ design for the zero coefficients and scaling factor, we observe that the distortion curve has a very small slope for VQ levels of 6 bits = 64 codewords and higher. Based on these observations we chose a 6-bit VQ for encoding the LSF coefficients, and another 6-bit VQ for encoding the zero coefficients and the scaling factor. These choices for the VQ are not final, and are further optimized based on WB-PESQ MOS performance, as explained later in this section. Each of the 2 past difference signals, and the 6 past reconstructed signals, were scalar quantized using 6-bit μ-law PCM derived from the G.711 encoder by ignoring the 2 least significant bits (LSBs). An 8-bit scalar quantizer was found to be adequate to accurately encode the integer pitch values. Since from Eqs.(2) and (3), only the signs of the two partially reconstructed signals are required at the decoder, they were encoded using 1 bit each.

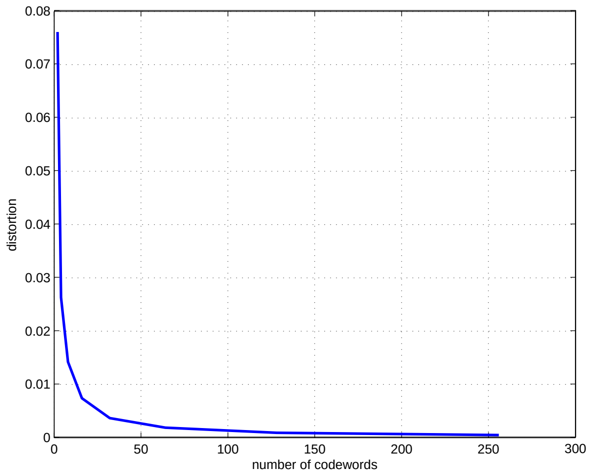


Figure 6. Codebook sizes and corresponding distortion in VQ design for LSF coefficients

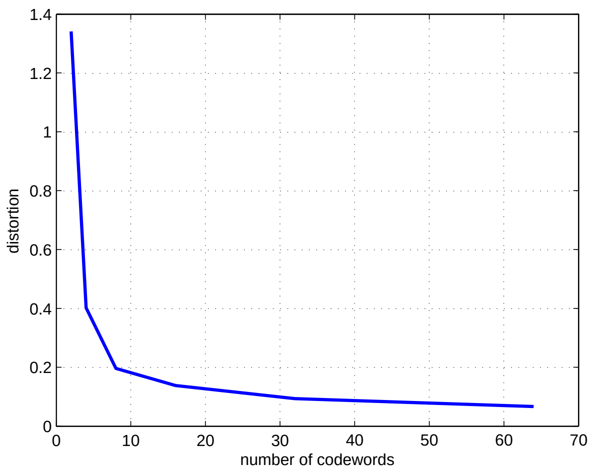


Figure 7. Codebook sizes and corresponding distortion in VQ design for zero coefficients and scaling factor

The bits alloted for encoding each parameter under the above described quantization method are specified in the second column in Table III. This quantization scheme encodes the side information using 70 bits. It results in a small quality degradation relative to the case when uncoded side information is available at the decoder, as demonstrated later in Fig. 11, while still performing significantly better than the *Modified Decoder Update* scheme. As a next step, we explored methods to reduce further the number of bits in the side information without compromising the voice quality. Using the quantization method described above as a basis, we evaluated the effect of varying the quantization levels used for the vector and scalar quantizers on WB-PESQ MOS performance.

TABLE III.
QUANTIZATION OF THE SIDE INFORMATION

Parameters	**Default (bits)**	**Updated (bits)**
VQ for pole LSF coefficients	6	6
VQ for zero coefficients and scaling factor	6	7
μ-law PCM coding for reconstructed and difference signals	48	32
Sign of partially reconstructed signals	2	2
Pitch information	8	8
Total	**70**	**55**

For encoding the 2 past difference signals and the 6 past reconstructed signals, a range of quantization levels from 6-bit μ-law PCM down to 2-bit μ-law PCM were considered. Starting with the 6-bit μ-law quantizer, the least significant bit (LSB) was successively dropped to obtain the coarser quantization levels, and the performance was evaluated. Default quantization levels (column 2 of Table III) are employed for encoding the remaining parameters in the side information. The reduction in the number of quantization bits/sample increases the quantization intervals but does not change the dynamic range of the μ-law PCM quantizer. The average WB-PESQ MOS values obtained for 100 different loss patterns, for a speech file containing male and female speech of duration 8 seconds, are plotted in Fig. 8. The 4-bit μ-law PCM is seen to provide a performance similar to 6-bit μ-law PCM and slightly better than the 3-bit μ-law PCM, and is therefore chosen to encode the past difference and reconstructed signals.

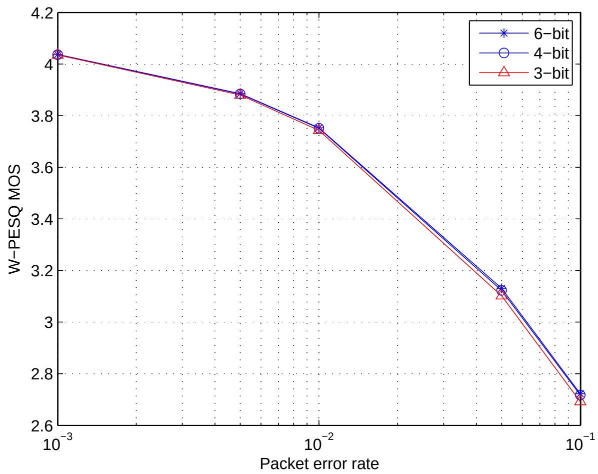


Figure 8. Performance of coded side information scheme under different quantization levels used for encoding the past reconstructed and difference signals

For the LSF coefficients, we compare the performances for 7-bit and 6-bit VQs respectively. The past reconstructed and difference signals were encoded using 4-bit μ-law PCM, while the remaining parameters were encoded using the default values (from column 2 in Table. III). The average WB-PESQ MOS values for these two cases, obtained for 100 different loss patterns for a speech file containing male and female speech of duration 8 seconds, are shown in Fig. 9. We observe that the performance of the 6-bit VQ is similar to the 7-bit VQ, and so we stick to the choice of using a 6-bit VQ for the LSF coefficients .

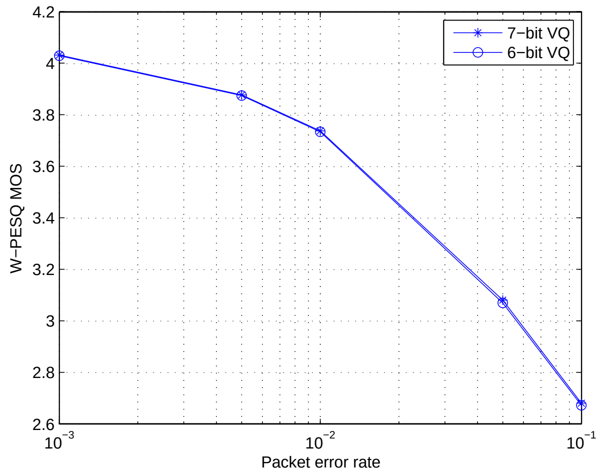


Figure 9. Performance of coded side information scheme under different quantization levels used for encoding the LSF coefficients

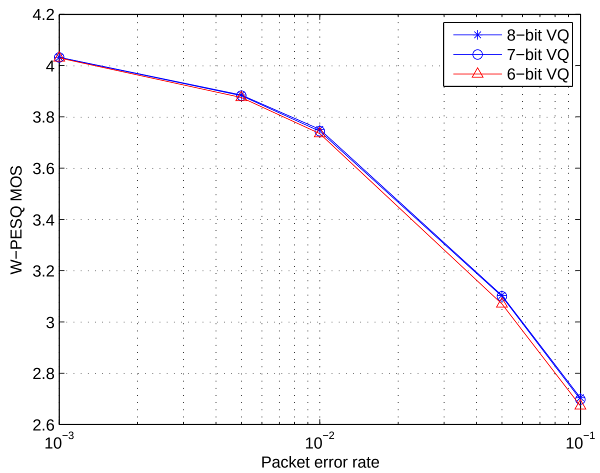


Figure 10. Performance of coded side information scheme under different quantization levels used for encoding the zero coefficients and the scaling factor

The 6 zero coefficients in Eq. (3) and the logarithmic scaling factor in Eq. (10) were jointly encoded using 6-bit, 7-bit and 8-bit VQs. As in the case of the LSF quantization, the past reconstructed and difference signals were encoded using 4-bit μ-law PCM, while the remaining parameters were encoded using the default values (from column 2 in Table. III). The average WB-PESQ MOS values for each of these quantization levels under different PLRs, obtained for 100 different loss patterns, for a speech file containing male and female speech of duration 8 seconds, are shown in Fig. 10. The 7-bit VQ offered an improvement in WB-PESQ of 0.03 at a PLR of 5% over the 6-bit VQ. However, no improvement in WB-PESQ is observed when comparing the 8-bit VQ with the 7-bit VQ. Therefore, we choose the 7-bit VQ to encode the zero coefficients and the scaling factor. Since the 6-bit VQ for the pole coefficients and the 7-bit VQ for the zero coefficients and the scaling factor generate only 13 bits of side information per frame, further efforts at reducing the number of bits used for the VQ quantizers were not pursued.

The updated set of quantization levels based on the tests conducted above, are shown in the last column of Table III. Notably, only the VQ for the zero-coefficients and the scaling factor, and the number of scalar quantizer bits used for encoding the past difference and reconstructed signals differ from those used in the default case. The updated quantization method reduces the number of bits required to encode the side information from 70 to 55. The WB-PESQ MOS values for the uncoded side information, coded side information using 70 bits and 55 bits respectively, and the modified decoder update scheme are shown in Fig. 11. The average MOS values are obtained for 500 different packet loss patterns corresponding to each PLR, and two sets of speech files each containing 8 seconds of speech and comprised of one male and one female sentence. In the case of the perfect and coded side information based schemes, we assume that the uncoded or coded side information is available at the decoder. We observe that the coded side information scheme using 55 bits has a performance close to the coded case with 70 bits, and results in a small degradation relative to the case when uncoded side information is available at the decoder. Subsequent references to coded side information in this paper, refer specifically to the updated quantization scheme employing 55 bits.

From Fig. 11, the use of coded side information results in a drop in MOS of 0.16 at a PLR of 10% when compared to the uncoded side information based PLC scheme. However, the use of coded side information still provides a significant improvement in performance over the modified decoder update scheme. At a PLR of 10%, the PLC scheme using coded side information provides an improvement of greater than 0.4 WB-PESQ over the modified decoder update scheme. That this improvement is perceptually significant is further verified by informal listening tests. Additionally, from Table III, we see that the side information of about 7 bytes constitutes less than 10% of the total frame size of 80 bytes corresponding to 10 ms of G.722 coded speech.

VI. PLC PERFORMANCE FOR G.722 USED IN VOICE OVER WIRELESS LAN

The performance of the proposed *Side Information based PLC scheme* of sending state and pitch information

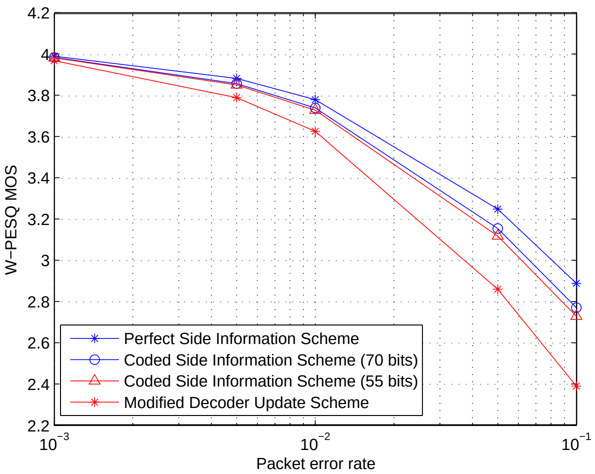
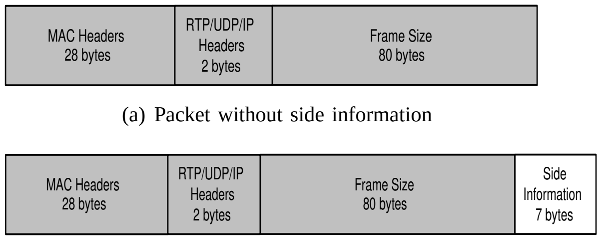


Figure 11. Performance comparison of the modified decoder update scheme and the side information scheme with perfect side information and coded side information using 70 bits and 55 bits

as side information is evaluated in a practical scenario involving voice communication over Wireless LANs. This is compared with the performance of the *Modified Decoder Update PLC* implemented for voice communication over WLANs.

The 802.11 packet format is comprised of the speech frame, an RTP/UDP/IP header of 40 bytes and a MAC header of 28 bytes. The use of Robust Header Compression (RoHC) can reduce the RTP/UDP/IP header size from 40 bytes down to 2 bytes [25]. We consider speech frames of duration 10 ms which corresponds to a speech frame size is 80 bytes for G.722 at 64 kbps. For the *Side Information based PLC*, the side information of 7 bytes is appended at the end of the speech frame in the packet. This side information is comprised of the pitch value of the previous frame and the state information needed to begin decoding that same frame. The packet formats for the *Modified Decoder Update PLC* and the *Side Information based PLC* are shown in Fig. 12. The addition of the side information of 7 bytes for the *Side Information based PLC* results in a small increase in packet size of less than 7% from 110 bytes to 117 bytes. In transmitting packets over the WLAN, no retransmissions were assumed. A random bit error rate (BER) channel model corresponding to an additive white Gaussian noise (AWGN) channel is considered. At the receiver, a packet is considered to be lost if any bit in the packet is in error.

The side information in each packet consists of the state information necessary for decoding the lower subband of the present frame and the pitch value of the previous frame. In Fig. 13, we demonstrate the concealment of the n^{th} frame and decoding of the $n+1^{th}$ frame when the n^{th} frame is lost, for the *Side Information based PLC*. In the event of the loss of the n^{th} packet, the decoder waits for the $n+1^{th}$ packet. This packet contains the $n+1^{th}$ frame and additionally side information that is comprised of the pitch value of the n^{th} frame and the



(a) Packet without side information

(b) Packet format with side information

Figure 12. Packet format for 802.11a voice packet using Robust Header Compression (RoHC) with and without side information

state information necessary for correct decoding of the $n+1^{th}$ frame. The pitch side information in the $n+1^{th}$ packet is employed in the basic pitch repetition scheme where pitch period repetitions of the previous correctly received frame ($n-1^{th}$ frame) are used to generate the concealment frame in place of the lost n^{th} frame. The state side information in the $n+1^{th}$ packet is then used to resynchronize the lower subband for proper decoding of the $n+1^{th}$ frame.

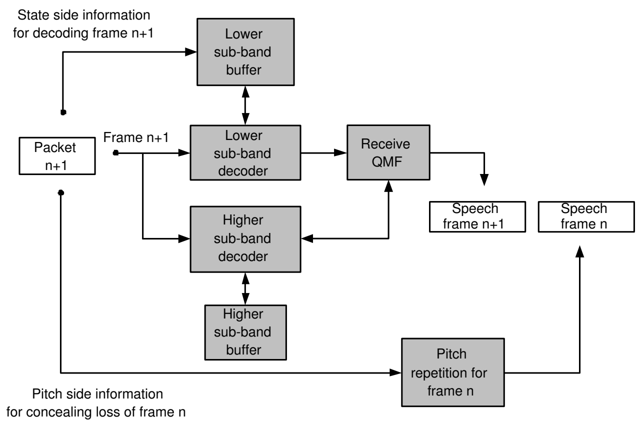


Figure 13. Decoding of the n^{th} and $n+1^{th}$ frame when the n^{th} frame is lost, for the side information based scheme

For a constant BER channel, the increase in packet size due to side information increases the PLR slightly; for example, a BER that results in a PLR of 10% for packets without side information, results in a PLR of 10.61% for packets with side information. For such a channel, the PLR values for packets without side information and the corresponding PLR values for packets with side information are presented in Table IV. The PESQ-MOS values associated with these sets of PLRs are shown in Fig. 14.

We observe from Fig. 14 that the side information based PLC scheme offers a significant improvement over the decoder update scheme. Comparing Fig. 14 with Fig. 11, we observe that for the side information based PLC scheme, the WB-PESQ undergoes a drop of less than 0.04 while going from a PLR of 10% to 10.61%. The increase in WB-PESQ MOS shown at a packet error rate of 0.05 is the difference between unacceptable and acceptable delivered voice quality. Thus, the increase in

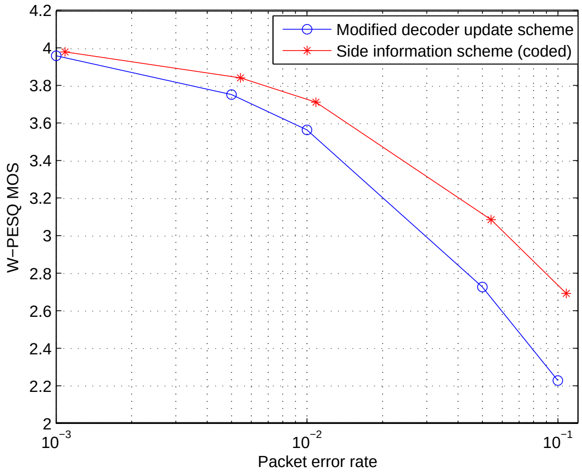


Figure 14. Performance comparison for packets with and without side information

TABLE IV.
CORRESPONDING PLRS FOR THE MODIFIED DECODER UPDATE AND THE CODED SIDE INFORMATION BASED SCHEMES IN A BER CHANNEL

Modified Decoder Update	Coded Side Information
0.1%	0.1064%
0.5%	0.532%
1%	1.063%
5%	5.31%
10%	10.61%

packet size due to the use of side information results in a small increase in PLR for a constant BER channel but causes a negligible degradation in speech quality. Thus, in a realistic scenario involving voice communication over wireless LANs, the use of the side information based PLC scheme in which the side information is appended with each packet before being transmitted results in a significant improvement in performance over the modified decoder update scheme, under a BER channel.

VII. CONCLUSIONS

We have proposed and investigated the performance of two new PLC schemes for the G.722 speech codec. While the experiments have been performed on the G.722 speech codec, the proposed schemes are clearly applicable to the G.726 speech codec as well. Our results show that by using a small amount of side information, a significant improvement in the PLC performance of the G.722 speech codec can be achieved. A method for quantizing the side information was implemented and the performance of the *Side Information based PLC* was evaluated for voice over wireless LANs. The side information in each frame increases the packet size by less than 7% and introduces an additional delay equivalent to a single frame. The proposed side information based PLC substantially improves the robustness of the G.722 coded speech to packet losses with minimal increase in bandwidth and delay requirements.

REFERENCES

[1] P. Mermelstein, "G.722, A New CCITT Coding Standard for Digital Transmission of Wideband Audio Signals," *IEEE Communications Magazine*, vol. 26, no. 1, pp. 8–15, January 1988.

[2] J. D. Gibson, T. Berger, T. Lookabaugh, and R. L. Baker, *Digital Compression for Multimedia*. Morgan Kauffman, San Francisco, 1998.

[3] R. V. Cox, "Speech coding standards," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Elsevier, 1995, pp. 49–78.

[4] Cisco, "Cisco IP Phone 7985G Data Sheet," 2005.

[5] 3Com, "3101 Basic IP phone Data Sheet," 2005.

[6] Grandstream, "BudgetTone-100 IP Phone Data Sheet," 2006.

[7] B. W. Wah, X. Su, and D. Lin, "A Survey of Error-Concealment Schemes for Real-Time Audio and Video Transmissions over the Internet," *International Symposium on Multimedia Software Engineering*, 2000.

[8] ITU-T Recommendation G.711 Appendix I, "A high quality low-complexity algorithm for packet loss concealment with G.711 ," Oct. 1999.

[9] 3GPP TS 26.090 V6.0.0, "Adaptive Multi-Rate (AMR) speech codec: Transcoding functions," Dec. 2004.

[10] R. Lefebvre, G. Philippe, and R. Salami, "A study of design compromises for speech coders in packet networks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 2004.

[11] P. Gournay, F. Rousseau, and R. Lefebvre, "Improved packet loss recovery using late frames for prediction-based speech coders," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 2003.

[12] S. Kubota, A. Dobashi, M. Suzuki, T. Hasumi, and S. Kato, "Improved ADPCM Voice Signal Transmission Employing Click-Noise Detection Scheme for TDMA-TDD Personal Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 1, pp. 108–113, February 1997.

[13] M. Serizawa and Y. Nozawa, "A Packet Loss Concealment Method using Pitch Waveform Repetition and Internal State update on the Decoded speech for the Sub-band ADPCM Wideband Speech Codec," *IEEE Speech Coding Workshop*, pp. 68–70, 2002.

[14] N. Shetty and J. D. Gibson, "Improving the Robustness of the G.722 Wideband Speech Codec to Packet Losses for Voice over WLANs," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006.

[15] Internet Link: http://www.itu.int/ITU−T/studygroups /com16/results.html, *accessed February 2, 2007*.

[16] J. Thyssen, R. Zopf, J.-H. Chen, and N. Shetty, "A Candidate for the ITU-T G.722 Packet Loss Concealment Standard," *to be presented at the IEEE International Conference on Acoustics, Speech and Signal Processing at Honolulu, Hawaii*, April 2007.

[17] ITU-T COM12-D7, "Proposed Modification to draft P.862 to allow PESQ to be used for quality assessment of wideband speech," Feb. 2001.

[18] ITU-T Recommendation P.862, "PESQ: An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Feb. 2001.

[19] C. Morioka, A. Kurashima, and A. Takahashi, "Proposal on Objective Speech Quality Assessment for Wideband Telephony," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2005.

[20] N. Kitawaki, N. Nagai, and T. Yamada, "Objective Quality Assessment of Wideband Speech Coding," *IEICE Transactions on Communication*, vol. E88-B, no. 3, pp. 1111–1118, March 2005.

[21] ITU-T Recommendation G.722, "7 khz audio coding within 64 kbps," Nov. 1988.

[22] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Prentice-Hall, New Jersey, 1984.

[23] ITU-T Recommendation P.341, "Characteristics of Wideband Terminals," 1994.

[24] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, 1992.

[25] C. Bormann, C. Burmeister, M. Degermark, H. Fukushima, H. Hannu, L.-E. Jonsson, R. Hakeberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, "Robust header compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed." *IETF RFC3095*, July 2001.

Niranjan Shetty is a Ph.D. candidate in Electrical and Computer Engineering at the University of California, Santa Barbara. He received his MS in Electrical and Computer Engineering from the University of California, Santa Barbara, in 2004. His research focus is on investigating methods to improve voice quality over heterogeneous multi-hop networks, and specifically addresses voice packet loss concealment and tandem performance of speech coders in these networks.

Jerry D. Gibson is Professor of Electrical and Computer Engineering and Professor in the Media Arts & Technology Graduate Program at the University of California, Santa Barbara. He is co-author of the books Digital Compression for Multimedia (Morgan-Kaufmann, 1998) and Introduction to Nonparametric Detection with Applications (Academic Press, 1975 and IEEE Press, 1995) and author of the textbook, Principles of Digital and Analog Communications (Prentice-Hall, second ed., 1993). He is Editor-in-Chief of The Mobile Communications Handbook (CRC Press, 2nd ed., 1999), Editor-in-Chief of The Communications Handbook (CRC Press, 2nd ed., 2002), and Editor of the book, Multimedia Communications: Directions and Innovations (Academic Press, 2000).

Dr. Gibson was Associate Editor for Speech Processing for the IEEE Transactions on Communications from 1981 to 1985 and Associate Editor for Communications for the IEEE Transactions on Information Theory from 1988- 1991. He was President of the IEEE Information Theory Society in 1996, and a member of the Speech Technical Committee of the IEEE Signal Processing Society from 1992-1994. Dr. Gibson served as Technical Program Chair of the 1999 IEEE Wireless Communications and Networking Conference, Technical Program Chair of the 1997 Asilomar Conference on Signals, Systems, and Computers, and General Co-Chair of the 1993 IEEE International Symposium on Information Theory. Currently, he serves on the Steering Committee for the Wireless Communications and Networking Conference, and he is an elected Member-at-Large on the Communications Society Board of Governors for 2005-2007. He serves on the Editorial Boards of the IEEE Transactions on Information Forensics and Security and the Hindawi Open Access journal Advances in Multimedia.

In 1990, Dr. Gibson received The Fredrick Emmons Terman Award from the American Society for Engineering Education, and in 1992, was elected Fellow of the IEEE for contributions to the theory and practice of adaptive prediction and speech waveform coding. He was co-recipient of the 1993 IEEE Signal Processing Society Senior Paper Award for the Speech Processing area.

His research interests include data, speech, image, and video compression, multimedia over networks, wireless communications, information theory, and digital signal processing.