IMPROVING THE ROBUSTNESS OF THE G.722 WIDEBAND SPEECH CODEC TO PACKET LOSSES FOR VOICE OVER WLANS

Niranjan Shetty and Jerry D. Gibson

Department of Electrical & Computer Engineering
University of California, Santa Barbara CA 93106
e-mail :{niranjan, gibson}@ece.ucsb.edu

ABSTRACT

Since the G.722 wideband speech codec offers higher quality and naturalness than G.711, is low in complexity, has low delay, and tandems well with other codecs, it is an attractive codec for Voice over IP and Voice over Wireless LANs. However, packet losses in G.722 not only require good concealment of the lost frame, but a lost frame results in a mismatch of the encoder/decoder states for the next correctly received frame following the lost frame. Although proprietary schemes exist, the G.722 codec has no standardized packet loss concealment (PLC) method. We present a new PLC method for G.722 and propose an efficient approach to sending the side information to resynchronize the encoder and decoder that greatly improves the robustness of the G.722 codec to packet losses.

1. INTRODUCTION

The G.722 speech codec decomposes the wideband speech input into two subbands, and uses backward Adaptive Differential Pulse Code Modulation (ADPCM) to code each subband. At the receiver, the coded speech in each subband is decoded using the ADPCM decoder, and reconstructed to give the decoded wideband speech. This simple coding and decoding operation causes a delay of only 3 ms, and provides good performance at rates of 48, 56 and 64 kbps [1]. 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 [2–4].

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 network impairments such as delays, 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. While proprietary schemes exist, the G.722 codec lacks a standardized PLC method.

A simplistic approach to concealing a frame loss is to use pitch period repetitions of a previous correctly received frame as the concealment frame. 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. A scheme for updating the internal state parameters was proposed by Serizawa and Nozawa [5], which involves updating the internal states of the ADPCM decoders by processing the concealment frame at the receiver. While this reduces the effect of clicks, their forgetting factor control tends to damp the gains of the decoded speech 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 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 an improved PLC performance. Additionally, the use of the exact pitch value of the lost frame in generating the concealment frame is seen to improve the PLC performance. These observations form the basis for the proposed scheme using quantized side information for improving the PLC performance of the G.722 codec.

The operation of the G.722 codec is provided in the next section. The modification to the decoder update scheme and the proposed side information based PLC scheme are described in Sections 3 and 4. Experimental results demonstrating the improvement obtained through the use of side information are presented in Section 5.

2. THE G.722 CODEC [6]

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.

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.

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

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

where $a_{i,n-1}$ and $b_{i,n-1}$ are the adaptive pole and zero coefficients, $d_t[n]$ 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:

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

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].

The zero coefficients are updated using the equation

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

for i=1,..,6.

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

$$\nabla(n) = \gamma \nabla(n-1) + W[I(n-1)] \quad (3)$$

where γ is the leakage constant and is 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 that described above for the encoder. The decoded output for each subband are then reconstructed using the receive QMF filters and summed to give the decoded wideband speech.

3. MODIFIED DECODER UPDATE SCHEME

In [5], the lost frame is first concealed by using pitch-period repetitions of the previous correctly received frame in a manner similar to the ITU PLC [7] scheme for G.711. For updating the state information, the concealment frame 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 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, it is observed that altering the forgetting factor for the adaptive quantizer damps the gain of the reconstructed signal and results in a worse performance than the pitch repetition scheme for certain packet loss patterns. In our implementation of the scheme, the quantization scaling factor is left unaltered, while the forgetting factors in Eq.(1) for updating the pole coefficients are altered to the values specified in [5]. The values for the forgetting factors are shown in Table 1. We refer to the scheme in [5] as the 'Decoder Update' scheme and our modified implementation as the 'Modified Decoder Update' scheme.

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

Table 1. Forgetting factor values for the different schemes

The modified decoder update scheme reduces the effect of clicks due to state mismatch between the encoder and decoder states following a lost frame but fails to eliminate it completely.

4. PERFECT SIDE INFORMATION BASED PLC

The side information associated with a frame is comprised of (1) the state information for the lower subband ADPCM decoder, needed for correctly decoding that frame as specified in Table 2 and (2) the pitch value of the previous frame. In the event of a packet loss, the decoder waits for the side information in the next correctly received packet. The proposed concealment method uses the accurate pitch value of the lost frame, extracted from the side information instead of estimating the pitch from the previous correctly received frame. Before decoding the frame in the correctly received packet following a loss, the lower subband ADPCM decoder state is updated using the side information available in the same packet.

$a_{i,n-1}$; $i = 1, 2$	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

Table 2. State information needed at the decoder for reconstructing sample at time instant n

We began our experiments by first studying the effect on PLC performance of providing uncoded state information to the ADPCM decoders. We then investigated the effect of using accurate pitch of the lost frame for concealment at the decoder. Based on the improvement observed in each of the above experiments, a method for quantizing the side information was developed and the performance of the scheme with coded side information was analyzed.

5. EXPERIMENTAL RESULTS

The experiments were performed using two wideband speech segments of male and female speech as input. Each segment is comprised of two sentences by the same speaker and is 8 seconds in duration. The default operating rate of 64 kbps is used for the G.722 codec. The packet loss rates (PLR) considered are 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 Wideband PESQ (W-PESQ) has been shown to reflect perceptual quality well for coding conditions using the same codec [8], and was used to assess the performance of each scheme. The mean WPESQ-MOS values have been plotted in the figures. Informal listening tests were conducted to corroborate the W-PESQ results.

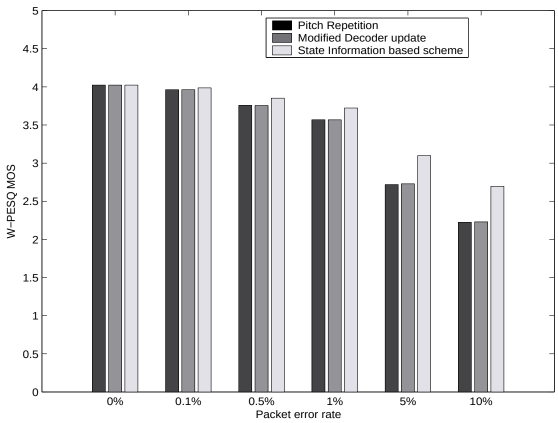

Fig. 1. Performance of G.722 PLC using 1) Pitch repetition 2) Modified decoder update 3) State side information

In Fig.1 we compare the performances of the pitch repetition scheme based on [7], the modified decoder update scheme described in Section 3 and the side information scheme which provides raw uncoded state information to each of the subband ADPCM decoders. We observe that the availability of perfect side information results in a significant improvement in the performance of the G.722 speech codec in the presence of packet losses, as compared to the pitch repetition scheme and our modified decoder update scheme.

The effect of providing state information to the ADPCM decoder in each subband individually, with the pitch estimated from the previous correctly received frame, is shown in Fig.2. We observe that communicating the state information for the lower subband only results in an improvement in performance comparable to that obtained when state information from both the subbands is available at the decoder. Based on this result, we send 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.

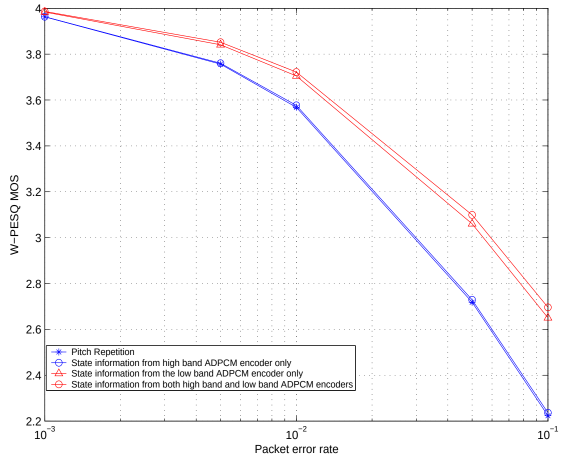

Fig. 2. Performance comparison of G.722 PLC with state information for individual subbands.

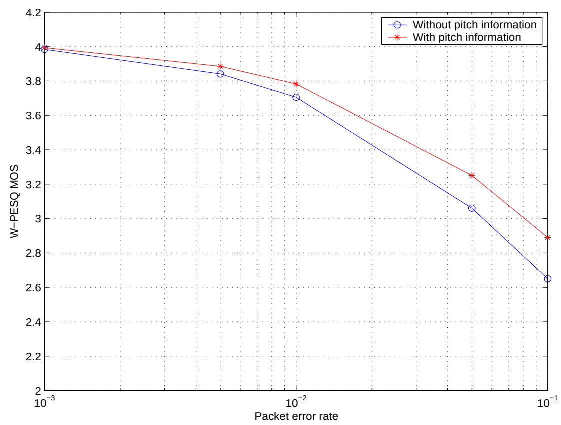

Fig. 3. Performance comparison of Side Information based scheme with and without pitch information.

The effect of using the exact pitch value of the lost frame for generating the concealment frame was analyzed next, and the performance of the side information based scheme with and without pitch information is shown in Fig.3. An improvement in performance is observed when pitch is included in the side information.

A method for quantizing the side information comprised of the state information specified in Table 2 and the pitch information, is investigated next. For the state information, the two pole coefficients in Eq.(1) are first converted into LSF coefficients to ensure a stable response after decoding at the decoder. The LSF coefficients were then encoded using a 6-bit VQ codebook designed using the LBG algorithm. The 6 zero coefficients in Eq.(2) and the logarithmic scaling factor

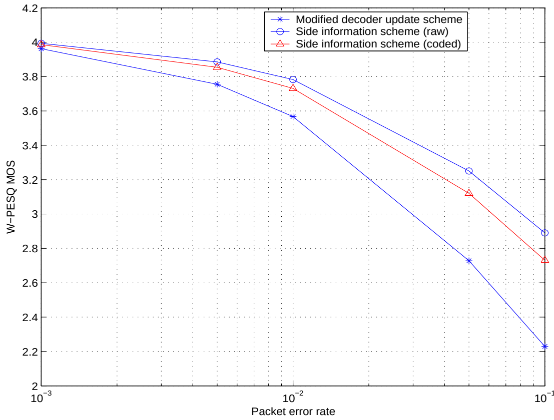

Fig. 4. Performance comparison of the modified decoder update scheme and the side information scheme with raw and coded side information

in Eq.(3) are jointly encoded using a 7-bit VQ. Since only the signs of the 2 partially reconstructed signals are required at the decoder, they can be losslessly encoded using 1 bit each. Each of the 2 past difference signals and the 6 past reconstructed signals are encoded using 4-bit μ-law PCM. The quantization scheme used for the different parameters in the side information are shown in Table 3.

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. For training the pole LSF coefficients, the pole coefficients in the state information were converted into LSF coefficients and were used for training the codebook. This resulting training data contained 128,000 LSF vectors. For the codebook design, uniform binary divisive clustering is used. The training data and VQ codebook design for the zero coefficients and scaling factor is obtained similarly. A 4-bit μ-log PCM scheme is used to code the reconstructed and difference signals, and is implemented using the G.711 8-bit μ-log PCM, by ignoring the 4 least significant bits (LSBs).

Parameters	**Bits**
6-bit VQ for pole LSF coefficients	6
7-bit VQ for zero coefficients and scaling factor	7
4-bit μ-log PCM coding for reconstructed and difference signals	32
Sign of partially reconstructed signals	2
Pitch information	8
Total	**55**

Table 3. Quantization of the Side Information

We observe from Fig. 4 that the use of coded side information results in a drop in MOS of 0.16 at a PLR of 10% when compared to the side information based PLC scheme using uncoded side information. 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 >0.4 WPESQ-MOS over the modified decoder update scheme. This is further verified by informal listening tests.

From Table 3, we observe that the side information constitutes less than 10% of the total frame size. In real-time communication over Wireless LANs, the total packet size with side information will be 28(MAC headers) + 2(RTP/UDP/IP headers with RoHC) + 80(frame size) + 7(side information) = 117 bytes, while the packet size without side information is 110 bytes. Thus, the addition of side information results in an increase in packet size of less than 6 %. For a constant BER channel, the increase in packet size due to side information increases the PLR slightly; for example, a PLR of 10% for packets without side information corresponds to a PLR of 10.61 % for packets with side information. Even after accounting for this increase in PLR, the PESQ-MOS of the side information based scheme is comparable to the plot in Fig. 4, with a drop in MOS of < 0.04 while going from a PLR of 10% to 10.61%.

6. CONCLUSIONS

We have proposed and investigated the performance of two 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. The side information in each frame constitutes less than 6 % of the packet size and introduces an additional delay equivalent to a single frame. Thus, the proposed side information based PLC improves the robustness of the G.722 coded speech to packet losses, with minimal increase in bandwidth and delay requirements.

7. REFERENCES

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

[2] Spectra Link, “Netlink System Overview,” 2005.

[3] Cisco, “Cisco IP Phone 7960G Data Sheet,” 2005.

[4] 3Com, “3101 Basic IP phone Data Sheet,” 2005.

[5] 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.

[6] ITU-T Recommendation G.722, “7 khz audio coding within 64 kbps,” Nov. 1988.

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

[8] A. Takahashi C. Morioka, A. Kurashima, “Proposal on Objective Speech Quality Assessment for Wideband Telephony,” *ICASSP 2005*.