A CANDIDATE FOR THE ITU-T G.722 PACKET LOSS CONCEALMENT STANDARD†

*Jes Thyssen, Robert Zopf, Juin-Hwey Chen, and Niranjan Shetty**

Broadcom Corporation, Irvine, California, USA

ABSTRACT

This paper presents a candidate for the ITU-T G.722 packet loss concealment standard. The algorithm is based on waveform extrapolation in the speech domain. The strong backward adaptive nature of G.722 makes the state update during lost frames a challenge. The paper presents methods to update the G.722 subband decoder state memory during packet loss. Furthermore, novel techniques to facilitate smooth transition after packet loss are described. Formal subjective test results indicate that the algorithm far exceeds all requirements in the ITU-T terms of reference. Most techniques are applicable to G.726 as well, and some are easily extendable to other speech coders with memory.

***Index Terms*—** G.722, ADPCM, packet loss concealment

1. INTRODUCTION

With the emergence of Voice over Internet Protocol (VoIP), legacy speech coders are seeing a need for Packet Loss Concealment (PLC) algorithms. Previously, ITU-T has standardized G.711 Appendix I [1], a PLC algorithm for G.711. Also, G.728 Annex I [2] specifies a PLC algorithm for G.728. However, both G.726 [3] (narrowband ADPCM) and G.722 [4] (wideband subband ADPCM) are still lacking standard PLC algorithms. At the request of ETSI for use in DECT, ITU-T Q.10/SG16 initiated the standardization of a PLC algorithm for G.722 [5]. This paper presents a candidate for the G.722 PLC standardization.

The paper is organized as follows. Section 2 provides an overview of the G.722 PLC requirements and Section 3 presents a brief overview of G.722. Structural considerations and alternatives are presented in Section 4, and the submitted candidate algorithm is described in Section 5 in detail. Experimental results are reported in Section 6, and conclusions are presented in Section 7.

2. ITU-T G.722 PLC REQUIREMENTS

The G.722 PLC Terms of Reference (ToR) are given in [6]. The general speech quality requirement specifies that the candidate algorithm must be "better than" G.722 with "minimum codeword substitution". This requirement must be met for random and bursty packet loss. The loss rates for clean speech are 3%, 6%, and 3% with concurrent 0.1% Residual Bit Errors (RBE). Additionally, 1% Packet Loss (PL) is tested without a requirement. For speech in background noise (both music at -25 dB SNR and office noise at -20 dB SNR) the loss rate is 3%, again both random and bursty. The "minimum codeword substitution" means that the lowest quantization levels for both the low-band and high-band decoders of G.722 are used for samples of lost packets. In practice, two versions were included, with and without decoder reset. The better of the two scores are reported as the G.722 PLC Reference in this paper.

The ToR specifies packet sizes of 10 ms and 20 ms with zero delay increase compared to G.722 in same packet configuration.

An additional computational complexity of 5 WMOPS and an additional memory usage of 2 kwords of RAM and 5 kwords of ROM compared to the standard G.722 decoder are allowed.

3. G.722 OVERVIEW

G.722 decomposes the 16 kHz input into an 8 kHz low-band and an 8 kHz high-band signal (see Figure 1). Each subband signal is encoded with ADPCM at 6 bits per sample for the low-band and 2 bits per sample for the high-band for a total of 8 bits per sample at 8 kHz and a bit-rate of 64 kbps. The low-band ADPCM utilizes embedded coding and it can be decoded at either 4, 5, or 6 bits per sample corresponding to a total bit-rate of 48, 56, or 64 kbps.

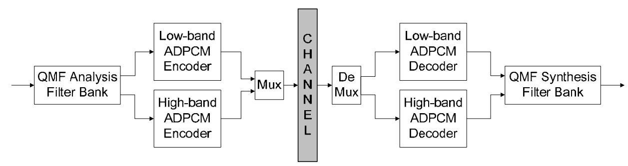

Figure 1: G.722 subband ADPCM.

Disregarding the details of the embedded capability of the low-band ADPCM, both ADPCM encoders fundamentally operate as shown in the left half of Figure 2, and the subband decoders as depicted in the right half of Figure 2. Note that the corresponding subband encoder and decoder have identical algorithms and signals for the backward adaptation of quantization scale factor, zero section of predictor, and pole section of predictor. Ideally, everything is updated in synchrony between encoder and decoder. However, in channel error conditions, temporary divergence is inevitable. As shown in Figure 2, the G.722 ADPCM subband coders utilize a backward adaptive pole-zero predictor with quantization of the prediction error based on backward adaptive scaling of the codebook. The backward adaptation is sample-by-sample resulting in potentially significant divergence of states for a packet size of 10 ms (80 samples in the subband domain).

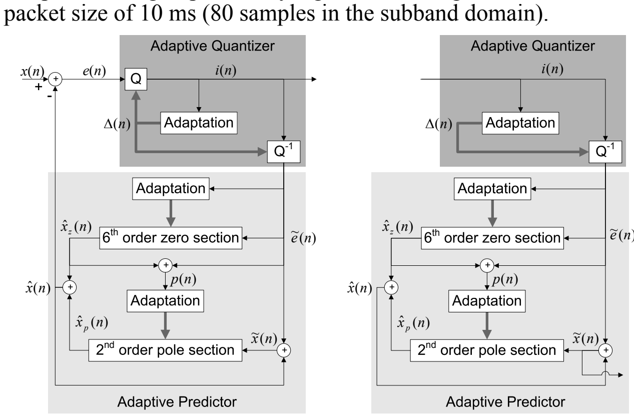

Figure 2: Left: ADPCM encoder, right: ADPCM decoder.

† This candidate became ITU-T G.722 Appendix III after submission of the original paper.
* Niranjan Shetty worked at Broadcom as a summer intern from University of California, Santa Barbara.

The encoder input is the speech signal, $x(n)$, and the output is the index, $i(n)$, while the decoder input is the index, $i(n)$, and the output is the reconstructed speech, $\tilde{x}(n)$. Note that "~" (tilde) indicates a reconstructed or quantized signal while "^" (hat) indicates a predicted signal. From a PLC point of view the decoder can be viewed as an excitation signal, $\tilde{e}(n)$, driving a pole-zero short-term synthesis filter.

4. STRUCTURAL CONSIDERATIONS

Designing a PLC algorithm for G.722 involves special consideration due to the split-band with independent low-band and high-band coding. Two design approaches come to mind:

1. Perform the PLC in the subband signal domain, or
2. Perform the PLC in the output signal domain.

Approach 1 has the advantage of possibly doing the PLC in the excitation domain of the subband decoders similar to CELP coders. The excitation is passed through the pole-zero short-term predictors to get the subband outputs, and subsequently through the QMF synthesis filterbank to get the output signal. State memory update would come naturally as the PLC excitation would directly replace the lost excitation and drive the two subband ADPCM decoders. Furthermore, PLC algorithms could be tailored specifically to the subbands placing more emphasis, quality and/or complexity wise, on the more important band and/or properties of a specific subband. A PLC synthesis based on a subband pitch extrapolated excitation signal, $\tilde{e}_{PLC}(n)$, is shown in Figure 3. The predictor configuration is identical to Figure 2 but slightly redrawn. Note that the backward adaptive quantizer is updated by approximating the prediction error signal, $e(n)$ with the extrapolated excitation signal, $\tilde{e}_{PLC}(n)$, and passing it through the quantizer and associated adaptation. Other meaningful configurations can be conceived along similar lines of thought.

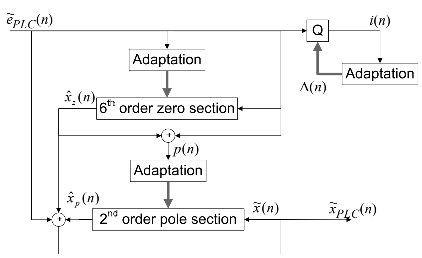

Figure 3: Subband ADPCM excitation domain PLC example.

One such configuration lets the PLC take place in the subband output signal domain using a NarrowBand (NB) PCM PLC such as G.711 Appendix I [1] and back-tracks from there to obtain reasonable state memory updates. The back-tracking could take place in the form of re-encoding similar to the suggestion in [7]. Such a configuration including both subbands is depicted in high level form in Figure 4, where $x_L(n)$ and $x_H(n)$ are passed to low-band and high-band ADPCM encoders, respectively.

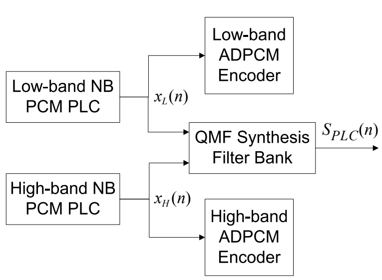

Figure 4: Approach 1 G.722 PLC.

Approach 2 has the advantage of performing the PLC in the output domain of the entire 8 kHz bandwidth in a single step, and not of the lower and upper 4 kHz separately. In particular, performing PLC of the upper 4 kHz during voiced speech can present a problem as a harmonic full-band signal does not necessarily produce a harmonic high-band signal. This would break the harmonic structure at the cross-over from the low-band to the high-band. However, performing the PLC in the output domain presents a greater challenge regarding update of the state memory of the subband ADPCM decoders. A solution is proposed in [7], which basically suggests re-encoding the WideBand (WB) PCM PLC output to update the state memory of the ADPCM subband decoders. This is depicted in Figure 5.

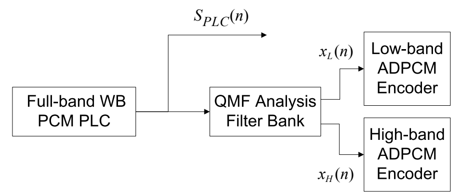

Figure 5: Approach 2 G.722 PLC.

During initial development, Approach 1 of Figure 4 and Approach 2 of Figure 5 were implemented using similar PCM PLCs configured for NB and WB, respectively. In the interest of saving complexity, experiments were also carried out with a greatly simplified high-band PCM PLC for Approach 1. However, expert listening indicated better perceptual speech quality using Approach 2, and hence, the submitted candidate is based on Approach 2. It performs periodic waveform extrapolation (PWE) of the WB speech waveform and update of state memory of the subband decoders by re-encoding. However, several techniques are introduced to the state memory update to improve quality, thereby deviating from pure re-encoding. Furthermore, several novel techniques were developed that greatly improve the transition from lost packets to received packets after a loss.

5. DETAILS OF CANDIDATE ALGORITHM

This section presents some of the novel algorithm techniques developed for the submitted Approach 2 G.722 PLC algorithm. The algorithm techniques are divided into 3 groups: (1) the WB PCM PLC, (2) those used to update the state memory of the G.722 decoder, and (3) those applied immediately following packet loss to reduce artifacts and distortion at the transition from lost packets to received packets.

It is important to note that due to the sample-by-sample adaptation of predictors and quantization scale factors of G.722, the parameters often exhibit modulation with local signal characteristics. As an example, for male voiced speech the low-band quantization scale factor will have a strong pitch modulation following the energy contour of the prediction error signal. This is shown in Figure 6 which contains the corresponding low-band input signal, prediction error signal, and quantization scale factor. Other parameters of G.722 exhibit a similar trend of modulation.

Note that all parameters and signals identified with a subscript "L" or "H" refer to the low-band or high-band ADPCM encoder/decoder, respectively.

5.1. Substituting the waveform during lost packets

The substitution of the speech waveform during packet loss (WB PCM PLC) is based on PWE of past speech and mixing of shaped noise dependent on the character of speech preceding the packet loss. For extended packet loss the output is gradually muted. A pitch period in the range of 2.5 ms to 16.56 ms is estimated, initially coarsely at 2 kHz sampling in the weighted speech domain and subsequently refined at 16 kHz sampling. The extrapolated waveform is extended beyond the current lost frame in order to: (1) accommodate the delay of the filterbank and provide sufficient samples to derive synchronized subband signals, (2) provide sufficient samples for overlap-add in the next frame, (3) facilitate time-warping and re-phasing described in Section 5.3.4. A block diagram of the WB PCM PLC is shown in Figure 7.

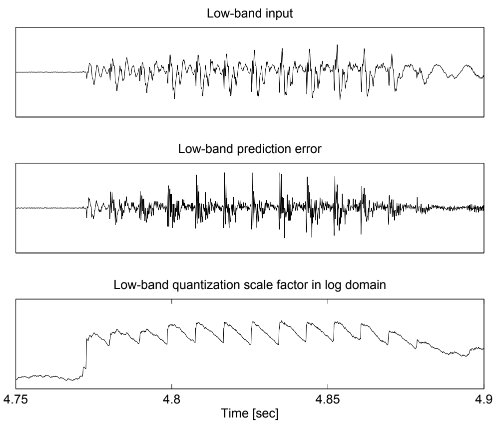

Figure 6: Example of pitch modulation of quantizer scale factor.

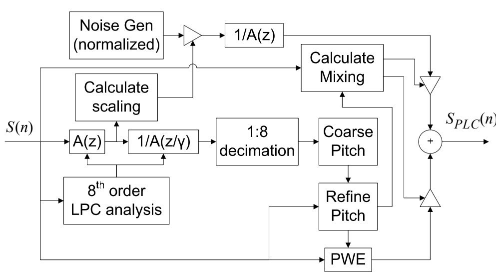

Figure 7: WB PCM PLC

5.2. Updating the G.722 decoder during lost packets

5.2.1 Modified re-encoding

The PLC output waveform is passed through the QMF analysis filter bank of G.722 in order to derive the low-band and high-band subband signals as shown in Figure 5. The subband signals are re-encoded by the respective encoders to update the decoder states. In order to save complexity an approximation is carried out. For the high-band ADPCM re-encoding it is recognized that the submitted algorithm does not use the quantization scale factor, $\Delta_H(n)$, at the first good frame after the packet loss, but instead resets the backward adaptation of the log scale factor to a running mean prior to the packet loss. Hence, if the prediction error signal is used unquantized, $\tilde{e}_H(n) = e_H(n)$, for the high-band adaptive predictor updates, the quantization of $e_H(n)$ can be saved. For the low-band ADPCM the scenario is slightly different. Due to the importance of maintaining the pitch modulation of the low-band adaptive quantization scale factor, $\Delta_L(n)$, it is advantageous to update this during packet loss. The low-band ADPCM encoder applies a 6-bit quantization of the difference signal, $e_L(n)$. However, only a subset of 8 of the quantization indices is used for updating the adaptive quantization scale factor. If the un-quantized prediction error signal is used for updating the adaptive low-band predictor, $\tilde{e}_L(n) = e_L(n)$, a less complex quantization can be deployed, yet maintaining identical update of the low-band adaptive quantization scale factor. The structures for modified low-band and high-band re-encoding are depicted in Figure 8.

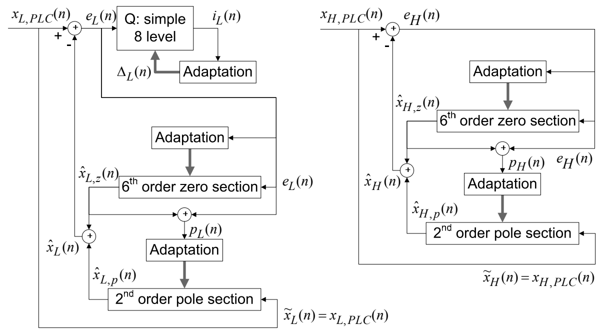

Figure 8: Modified re-encoding. Left: low-band, right: high-band.

5.2.2 Adaptive decoder reset

For long packet loss, it is sometimes better to reset the subband ADPCM decoders than to continue re-encoding. Intuitively, the divergence from the adaptation of the encoder increases with the duration of re-encoding with PLC output, and hence the probability of artifacts due to mis-tracking of signals or parameters increases. For those reasons both subband ADPCM decoders are reset after 60 ms of packet loss. However, earlier reset anywhere from 30 ms to 60 ms may occur depending on the monitoring of characteristics of the signal controlling the adaptation of the pole sections, $p_L(n)$ and $p_H(n)$, respectively. Notably, the two subband decoders are reset individually if their respective $p(n)$ signal is predominantly of constant sign. Such a signal fed to the sign-based LMS adaptation could drive the pole section to the stability limit, potentially causing a significant artifact after the packet loss due to an inappropriate state of the pole section.

5.3. Transition to received packets

5.3.1 High-frequency chirping

The signal decoded from good packets immediately following packet loss often contains very noticeable and annoying high-frequency chirping. This is particularly pronounced during silence. This originates from the high-band decoder, and is due to divergence of internal signals of the decoder. Many end up with almost constant sign in those cases. As described above, a $p_H(n)$ with constant sign can cause the high-band pole section of the predictor to drift towards the stability limit with a high gain due to the LMS adaptation being sign based. Adding DC removal to select internal signals helps "stabilize" the LMS adaptation. It was found that replacing $p_H(n)$ and $\tilde{x}_H(n)$ with high-pass filtered versions, $p_{H,HP}(n)$ and $\tilde{x}_{H,HP}(n)$, for the first 40 ms after packet loss eliminated the chirping:

$$\begin{aligned} p_{H,HP}(n) &= 0.97 \cdot [p_H(n) - p_H(n-1) + p_{H,HP}(n-1)] \\ \tilde{x}_{H,HP}(n) &= 0.97 \cdot [\tilde{x}_H(n) - \tilde{x}_H(n-1) + \tilde{x}_{H,HP}(n-1)] \end{aligned}.$$

This corresponds to a 3 dB cut-off of about 40 Hz.

5.3.2 *Pole section safety margin*

G.722 imposes the following constraint on the 2 pole section coefficients in order to ensure stability:

$$\begin{aligned} |a_2(n)| &\leq 0.75 \\ |a_1(n)| &\leq 1 - 1/16 - a_2(n) \end{aligned}.$$

As shown in [8], this is a triangular area where the constraint on $a_1(n)$ ensures a margin of $1/16$ to stability. Accordingly, an entity called "safety margin" is defined as

$$\alpha = 1 - |a_1(n)| - a_2(n),$$

where the regular G.722 constraint enforces a minimum safety margin of $\alpha_{\min} = 1/16$.

The pole section of the low-band ADPCM decoder often causes abnormal energy increase, perceived as a pop, after packet loss. Adaptively enforcing a more stringent constraint on the pole section of the adaptive predictor of the ADPCM low-band decoder greatly reduces this abnormal energy increase after packet loss. After packet loss, an increased minimum safety margin in $\alpha_{\min} = 3/16$ is enforced. It is gradually reduced to the standard minimum safety margin of G.722. Furthermore, a running mean of the safety margin prior to the packet loss is monitored and taken into account in the following way: the increased minimum safety margin at the first good frame(s) must not exceed the running mean of the safety margin prior to the packet loss.

5.3.3 *Adaptive resetting of quantizer scale factor*
For both the low-band and high-band ADPCM decoders it is beneficial to the performance in background noise to reset the adaptation of the quantizer scale factors to a running mean prior to the packet loss. This reduces energy drops otherwise often seen after packet loss in segments of background noise only. However, particularly for the low-band ADPCM decoder, this was found to occasionally produce large unnatural energy increases in voiced speech. Blindly resetting to a running mean will almost certainly cause miss-tracking in relation to pitch modulation as shown in Figure 6. On the other hand, re-encoding will result in a similar in-phase pitch modulation if the pitch period used by the WB PCM PLC is proper. Hence, the low-band quantizer scale factor is adaptively reset to a linear combination of the running mean and the re-encoded value based its character prior to the packet loss.

5.3.4 *Time-warping and re-phasing*
At the resumption of good packets there is no guarantee that the extrapolated waveform and the decoded waveform will be aligned properly, and objectionable artifacts may result. Unfortunately, since no delay is allowed, it is not possible to wait for received packets and adjust the WB PCM PLC to align the two waveforms. Instead, time-warping is devised. It will stretch or compress the time-scale of the signal in the first good frame to align the decoded waveform with the extrapolated waveform.

Additionally, the speech quality after packet loss will depend on proper phase of the subband ADPCM re-encoding. If the WB PCM PLC reconstructs synchronized speech, then it would leave the subband ADPCM decoders internal signals in suitable phase. However, due to the quick sample-by-sample adaptation of the backward adaptive parameters of ADPCM, even if it is off by only a few samples, it can result in large artifacts. The most sensitive is the low-band quantizer scale factor which, as shown in Figure 6, exhibits significant pitch modulation. This is addressed by a "re-phasing" technique that stops re-encoding of the WB PCM PLC output at a phase that it estimates to match the phase of the resuming packets.

Time-warping and re-phasing will be presented in detail and completeness in an upcoming paper [9].

6. EXPERIMENTAL RESULTS

The submitted G.722 PLC algorithm was tested by an independent subjective listening lab with speech material and error patterns not used during development. The G.722 reference and candidate algorithms ran at 64 kbit/s, and the G.729.1 [10] operated at 32 kbit/s. All listening tests were conducted with 32 listeners, and the speech material in American English included 3 female and 3 male talkers with 4 sentence pairs per talker. Table 1 includes the ACR MOS (clean speech) and DCR DMOS (speech in background noise) scores for 4 experiments. For both random and bursty packet loss, clean speech and speech in background noise/music, formal statistical analyses show that the submitted G.722 PLC candidate far exceeds all the requirements. Furthermore, it is interesting to note that the performance under packet loss of the submitted candidate is very close to that of G.729.1. In fact, the rate of degradation as function of loss rate is smaller or similar, indicating a robust PLC algorithm.

Table 1: MOS and DMOS.

Error Cond.	Exp	G.729.1	G.722PLC Candidate	G.722PLC Reference
None	Clean speech. Random packet loss	4.29	4.21	
1%		4.18	4.16	3.65
3%		4.00	3.99	2.80
6%		3.43	3.68	2.01
3%+0.1% RBE		NA	2.99	2.32
None	Clean speech. Bursty packet loss	4.16	4.11	
1%		4.03	3.90	3.85
3%		3.66	3.69	3.43
6%		3.15	3.24	2.87
3%+0.1% RBE		NA	3.13	2.90
None	Speech in 25 dB music	4.78	4.64	
3% random		4.20	4.31	2.66
3% bursty		4.10	4.04	3.56
None	Speech in 20 dB office	4.65	4.34	
3% random		4.28	4.12	2.29
3% bursty		3.99	3.77	3.04

7. CONCLUSION

This paper discussed several alternative structures for a G.722 PLC and presented solutions to handle challenges posed by G.722 in relation to PLC. The important features of a G.722 PLC algorithm submitted to the ITU-T standardization were presented. In formal subjective evaluations the candidate was shown to far exceed all speech quality requirements of the ITU-T ToR.

8. REFERENCES

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

[2] ITU-T Recommendation G.728 Annex I, "Frame or packet loss concealment for the LD-CELP decoder," May 1999.

[3] ITU-T Recommendation G.726, "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)," October 1990 (Originally approved as G.721 and G.723, November 1988).

[4] ITU-T Recommendation G.722, "7 kHz audio-coding within 64 kbit/s," November 1988.

[5] P. Barrett and C. Lamblin, "Report of the third meeting of Study Group 16 (Geneva, 3 – 13 April 2006) - Working Party 3/16 (Media Coding)," ITU-T SG16, Document TD 235 R1 (PLEN/16), Geneva, April 3-13, 2006.

[6] C. Lamblin and H. Taddei, "Terms of Reference (ToR) and time schedule for the G.722 Packet Loss Concealment (G.722 PLC) standardisation," ITU-T WP3/16, Document AC-06-14, Geneva, June 6-9, 2006.

[7] 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, Ibaraki, Japan, October 6-9, 2002, pp 68-70.

[8] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[9] R. Zopf, J. Thyssen, and J.-H. Chen, "Time-warping and Re-phasing in Packet Loss Concealment," to be published.

[10] ITU-T Recommendation G.729.1, "G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729," May 2006.