A LOW COMPLEXITY PACKET LOSS CONCEALMENT ALGORITHM FOR ITU-T G.722

Balázs KÖVESI and Stéphane RAGOT

France Telecom, R&D Division/TECH/SSTP
2, Av. Pierre Marzin, 22307 Lannion, France
Email: {balazs.kovesi,stephane.ragot}@orange-ftgroup.com

ABSTRACT

This article presents ITU-T G.722 Appendix IV which is a Packet Loss Concealment (PLC) algorithm recently standardized by ITU-T for G.722 decoding in the presence of frame erasures. This algorithm is suitable for applications that may encounter frame erasures or packet losses with a special focus on complexity constraints For example, G.722 Appendix IV is very suitable for DECT Next Generation and VoIP using low cost devices. Besides, we also describe some minor algorithmic modifications to G.722 Appendix IV that improve subjective quality. We discuss G.722-PLC performance based on formal ITU-T test results as well as additional informal experiments.

***Index Terms*—** Speech coding, audio coding, error correction, standardization

1. INTRODUCTION

Though one of the oldest speech and audio coding standards (dated 1988), ITU-T G.722 [1] is widely used and even planned to be further integrated in new telecommunication services such as the New Generation of Digital Enhanced Cordless Telecommunications (DECT) also called CAT-iq. Indeed G.722 has many attractive features, such as high wideband (50-7000 Hz) quality at 56 and 64 kbit/s, bitstream scalability from 48 to 64 kbit/s, low complexity, low delay and no IPR (Intellectual Property Rights) cost. However, until recently, the G.722 standard did not include any packet loss concealment algorithm. Following a request by ETSI TC DECT, ITU-T SG16 has launched the standardization of G.722-PLC. Three organizations participated in the fixed-point G.722-PLC selection phase: Broadcom, France Telecom, and Oki. The G.722PLC standardization process ended in Nov. 2006. Eventually, ITU-T approved two G.722-PLC algorithms: G.722 Appendixes III and IV. Formal selection test results demonstrated that both Appendixes meet the same quality requirements. In addition they showed that Appendix III has higher quality but increases G.722 decoding complexity by about 87%, while Appendix IV brings almost no additional complexity compared with G.722 normal decoding (about +2%). Thus, these two appendixes offer different quality-complexity trade-offs.

The objective of this article is to present G.722 Appendix IV, which has been developed by France Telecom. We explain how this algorithm has been designed. In particular, during the development of G.722 Appendix IV the main goal was to produce good quality in presence of frame erasures without increasing G.722 decoding complexity and minimizing extra ROM, in order to facilitate its integration in low capacity devices. Note that a complete description of G.722 Appendix IV can be found in [4].
After reviewing the background of G.722 PLC in Section 2, the functional description of G.722 Appendix IV is given in Section 3. Section 4 describes some minor modifications of G.722 Appendix IV that can improve subjective quality at the price of complexity increase. Complexity and quality results comparing G.722 Appendixes III and IV and the modified Appendix IV are given in Section 5, followed by the conclusion.

2. BACKGROUND OF G.722-PLC

2.1. G.722 decoding

The G.722 decoder comprises 3 main stages: low band (0-4 kHz) embedded ADPCM (Adaptive Differential Pulse Code Modulation) decoding (6, 5 or 4 bits per sample), high-band ADPCM decoding (2 bits per sample) and Quadrature Mirror Filter (QMF) synthesis.
ADPCM coding is highly recursive: in both bands the quantization scale factor, 6 Moving Average (MA) prediction coefficients and 2 AutoRegressive (AR) prediction coefficients are updated sample by sample. In the absence of transmission errors this update is performed in a synchronous way using the same quantized difference and reconstructed signals at both encoder and decoder. In case of erased frames the synchronization is lost, special care is needed to avoid artifacts and help resynchronization.

2.2. Main requirements of G.722-PLC selection

G.722-PLC candidates were tested in 4 experiments (see Table 1). Note that G.722 operated with 10 ms frames.

Table 1: Selection test experiments (FER: Frame Error Rate, RBER: Random Bit Error Rate).

Exp1a: Clean speech random	1-3-6% Random FER 3% Random FER + 0.1% RBER
Exp1b: Clean speech bursty	1-3-6% Bursty FER 3% Bursty FER + 0.1% RBER
Exp2a: Speech in back. music	Random and bursty FER at 3%,
Exp2b: Speech in office noise	Random and bursty FER at 3%

A trivial solution to handle frame erasures in G.722 is to fill out the missing parts of the bitstream by the codeword that corresponds to the minimal decoded value (0xFF). This solution was named PLC0. The requirements for all 4 experiments were to be better than the PLC0 reference [5].

3. DESCRIPTION OF G.722 APPENDIX IV

Frame erasures are indicated to the decoder through the Bad Frame Indication (BFI). In the absence of frame erasures G.722 Appendix IV is identical to G.722 decoding, except the decoded low- and high-band signals are memorized. In case of bad frame, in low band, after analyzing the past low-band synthesis, the missing signal is extrapolated using Linear-Predictive Coding (LPC), pitch synchronous period repetition and adaptive muting ; once a good frame is received after an erased frame the low-band decoded signal is cross-faded with the extrapolated signal. In high band, the decoder repeats pitch synchronously the previous frame with adaptive muting and high-pass post-processing. The ADPCM states are updated after each frame erasure.

3.1. Low-band frame extrapolation

The extrapolation of missing frames in the lower band is illustrated in Figure 1. It comprises an analysis part followed by a synthesis of the signal $yl(n)$, $n = 0,\dots,L-1$ where L is the length of the missing signal (L=80 for 10 ms frames). The analysis is made on the memorized past valid signal $zl(n)$, $n = -297,\dots,-1$. The length of this buffer (297 samples) allows to store at least the last two pitch periods.

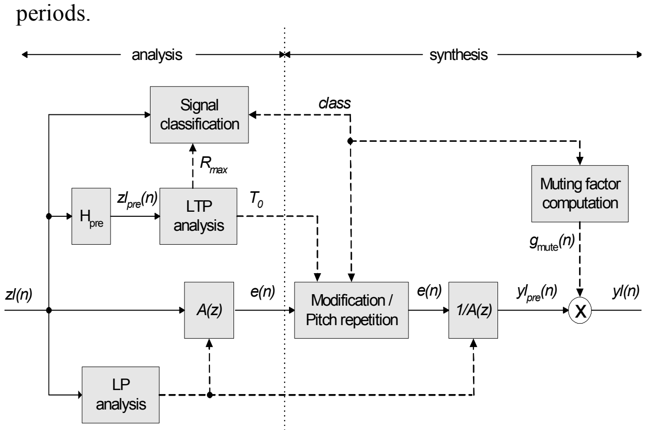

Figure 1: Low band extrapolation of missing frames.

3.1.1. LP analysis

The 8th order Linear Prediction (LP) analysis is made on the last 80 past lower band output samples. After windowing by an asymmetrical (70+10 samples) Hamming window, the autocorrelation function is computed including 60 Hz bandwidth expansion and 40 dB white noise correction, followed by the Levinson-Durbin algorithm.

The past signal $zl(n)$, $n = -297,\dots,-1$ is filtered through the obtained $A(z)$ filter to produce the LP residual signal $e(n)$, $n = -289,\dots,-1$.

3.1.2. LTP analysis

The Long-Term Prediction (LTP) analysis is preceded by a 50 Hz cut-off frequency high-pass pre-processing filter to remove undesired low-frequency components. Then the pitch period delay T_0 is determined using the past valid pre-processed signal $zl_{pre}(n)$, $n = -288,\dots,-1$. First this signal is decimated by a factor 4 and then filtered by a 2nd order weighting filter $B(z/0.94)$ where the coefficients of $B(z)$ are obtained by 2nd-order LP analysis of the decimated signal.

A first open loop pitch delay estimation on this weighted decimated signal T_{ds} is then refined in the preprocessed signal domain by searching the cross-correlation maximum in the neighborhood of $T=4T_{ds}$ to obtain T_0.

3.1.3. Signal classification

The PLC strategy depends on the past output signal characteristics. Using features (normalized correlation, lower- and higher band ADPCM scale factor ratios, zero crossing rate) and peak detection, the signal preceding an erasure is classified into one of the five following classes:

- TRANSIENT for transients with large energy variation (e.g. plosives)
- UNVOICED for unvoiced signals
- VUV_TRANSITION corresponding to a transition between voiced and unvoiced signals
- WEAKLY_VOICED for weakly voiced signals (e.g. onset or offset of vowels)
- VOICED for voiced signals (e.g. steady vowels)

3.1.4. LP residual modification

The extrapolated signal is the output of the LP filter $1/A(z)$ excited by an excitation signal $e(n)$, $n=0,\dots L-1$, which is the pitch synchronous repetition of the past LP residual signal.

Before performing the pitch repetition procedure, the last pitch period $e(n)$, $n=-T,\dots-1$ is modified if the signal class is WEAKLY_VOICED or VUV_TRANSITION. The modification consists in limiting the magnitude of each sample in the last pitch period as follows:

$$e(n) = \min\left(\max_{i=-2,\cdots,+2}\left(\left|e\left(n-T_0+i\right)\right|\right),\left|e\left(n\right)\right|\right)\times sign\left(e(n)\right),\quad n=-T_0,\cdots,-1$$

3.1.5. Pitch repetition of LP residual

If the signal class is VOICED, the excitation signal $e(n)$ is obtained by pitch-synchronously repetition:

$$e(n) = e\left(n-T_0\right),\quad n = 0,\cdots,L-1$$

If the signal class is not VOICED, this pitch-synchronous repetition is modified to avoid over-voicing by introducing a sample by sample jitter in which samples are swapped two by two:

$$e(n) = e\left(n-T_0+(-1)^n\right),\quad n = 0,\cdots,L-1$$

As this procedure is more efficient for odd pitch values, if T_0 is even, it is increased by 1. Note that 10 ms extra samples are also generated to prepare the signal for cross-fading.

3.1.6. Case of successive erased frames

In the case of a bad frame following a bad frame, the analysis parameters ($A(z)$, T_0, signal class) computed for the first erased frame are kept. The excitation signal generated as in Section *3.1.5* is filtered by the LP filter $1/A(z)$ to produce the extrapolated signal.

3.1.7. Adaptive muting

Both lower-band and higher-band extrapolated signals are muted before the QMF synthesis. The adaptive muting factor $g_{mute}(n)$ is updated sample by sample. The decreasing speed depends on the signal class as illustrated in Figure 2.

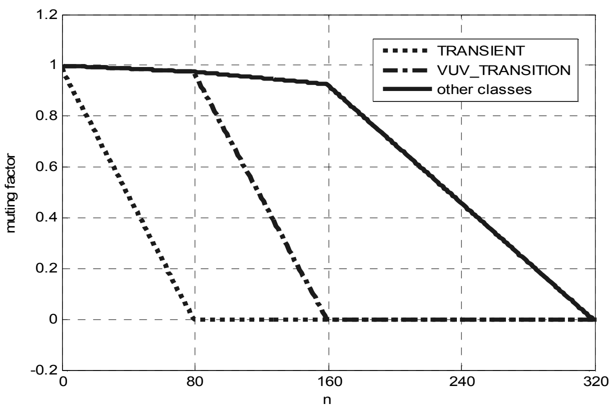

Figure 2: Muting factor as a function of sample index.

3.1.8. Lower-band ADPCM decoder states update

As mentioned in Section 2.1, in case of frame erasure the ADPCM decoder states can not be updated synchronously with the encoder states. This can cause serious artifacts when the decoding recovers after an erased frame. It is therefore important to properly update states. One solution could be to re-encode extrapolated samples which increases significantly decoding complexity. In G.722 Appendix IV a low complexity solution was chosen. The ADPCM states are updated once after an erasure in the following way (see [1] for details on G.722 states):

- The past quantized difference signal (DLT_i) is set to 0.
- The past reconstructed signal (RLT_i) is set equal to the last extrapolated samples.
- The past partially reconstructed signal (PLT_i) is set to half of the past reconstructed signal.
- The predictor output (SL) is set equal to the first future extrapolated sample. This sample is available as needed for the cross-fading too (see Section *3.1.9*).
- The zero section output (SZL) is set to half of the first future extrapolated sample.

Furthermore the quantizer scale parameters (NBL, DETL) are set to their initial value after 20 ms erasure: NBL = 0, DETL = 32.
Other low-band states are kept unchanged.

3.1.9. Cross-fading

In spite of the state memory update, cross-fading was also necessary to insure smooth transition between the extrapolated samples and the first valid decoded samples. It was observed that the first 20 decoded samples (2.5 ms) after an erasure are often unstable, but after this period the convergence of state variables is good enough to use output samples of "normal" G.722 decoding. Then, cross-fade is performed such that the weight of the first 20 decoded samples is 0, and then it increases linearly up to 1 during the next 60 samples while the weight of the corresponding extrapolated sample is complementary to 1.

3.2. High band frame extrapolation

Human hearing is less sensitive to high-band artifacts. That is why the high band extrapolation of G.722 Appendix IV is much simpler than in low-band. It simply consists in repeating pitch synchronously the past high-band output signal $zh(n)$ if the signal class is VOICED, otherwise the pitch period is set to 80 samples (10 ms). Then each extrapolated sample is muted by the same muting factor as the corresponding low-band sample (see Section *3.1.7*).

3.2.1. Highpass post-processing

In case of frame erasures, a DC offset of very small magnitude may appear in the high-band reconstruction and affect the first consecutive good frames. After QMF synthesis, this offset becomes an 8 kHz component. To avoid this annoying high-frequency noise, a first-order high-pass filter with a cut-off frequency of 50 Hz is used during the erased frames and the first 4 s following the erasure.

3.2.2. Higher-band ADPCM decoder state update

The high-band ADPCM decoder state update is simpler than in low band. After 10 ms erasure the logarithmic quantizer scale factor (NBH) is divided by two and the quantizer scale factor (DETH) is recomputed. After 20 ms erasure these variables are set their initial value (NBH = 0, DETH = 8). To save complexity, cross-fading is not applied in the high band.

3.3. QMF synthesis filterbank

The states of the QMF synthesis are updated automatically, as this filterbank works in a continuous way either with valid or extrapolated signals.

4. IMPROVEMENT OF G.722 APPENDIX IV

In G.722 Appendix IV the worst case complexity of good frame processing is about 10% higher than the worst case complexity of bad frame processing. This means that it is possible to add some extra operations in the bad frame processing part that may further improve the quality without increasing worst-case complexity. The following set of minor modifications were retained and tested:

- A single higher-precision 50 Hz high pass filter is used for the pre-processing in lower band analysis (see Section *3.1.2*) and the post processing in higher band (see Section *3.2.1*). Furthermore LPC analysis and classification are made on the pre-processed signal.
- Tuning of open-loop pitch estimation:
  - The best two down-sampled domain candidates are checked for voiced signals.
  - Extra verification are added to avoid two glottal pulses in one pitch period (in case the pitch period evolves slightly)
  - Inclusion of zero-crossing verification.
  - Limited pitch period length for TRANSIENT class.
- The repetition period of the excitation signal $e(n)$ and repetition period of the higher band signal are smoothed if the class is UNVOICED. Samples with amplitude higher then m_e are divided by 4, where m_e is 2.5 times the mean amplitude of the corresponding last 10 ms signal.
- After each erased frame the LPC synthesis filter $1/A(z)$ is weighted in the form of $1/A(z/0.99)$.
- The ARMA filter coefficients stored in the lower bad ADPCM state memory are weighted by γ=0.97 after each erased frame.
- Slower adaptive muting: The last linear part of the muting factor decrease function (see Figure 2) is changed in a way that the complete muting is achieved after 60 ms for VOICED,

WEAKLY_VOICED and UNVOICED classes and after 30 ms for the VOICED_UNVOICED class.

With these modifications the worst-case complexity of bad frame processing is still lower then that of good frame processing, the overall worst-case complexity is unchanged.

5. SUBJECTIVE TEST RESULTS

5.1. Selection test results

First the summary of the selection test results are given on Figure 3 and Figure 4 (from [6]). On these figures PLC A and PLC C correspond to the Appendixes III and IV respectively, the results are given using the MOS (Mean Opinion Score) scale. One can observe that both selected candidates are far better than the reference PLC0. On the other hand the difference between the two selected candidates is quite limited, 0.06 MOS in clean speech experiments and 0.16 MOS for speech mixed with background noise; according to the G.722-PLC global analysis [7].

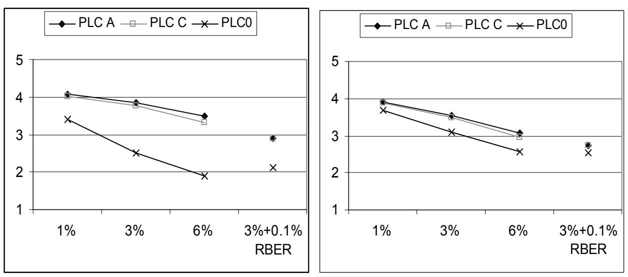

Figure 3: Selection test results for Exp1a and Exp1b.

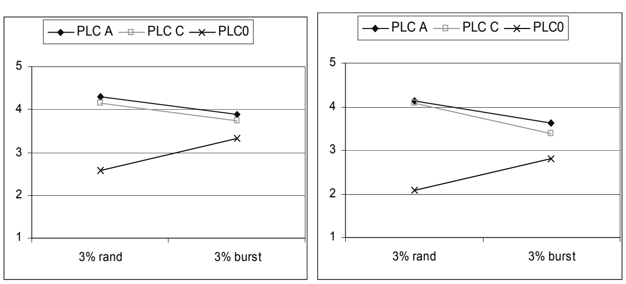

Figure 4: Selection test results for Exp2a and Exp2b.

5.1.1. Complexity analysis

While providing quite similar performance, G.722 Appendix IV has 46% less worst case complexity, 40% less average complexity, 69% less additional program ROM and 83% less additional data ROM than G.722 Appendix III.

5.2. Testing of proposed modifications to G.722 Appendix IV

Subjective test on selection test experiences is conducted to check the effect of the modifications presented in Section 4. The ITU selection test plan [8] was slightly modified to be able to compare Appendixes III, IV and the modified Appendix IV. Each experiment was performed by 12 expert listeners (3 per group). The MOS results can be seen on Figure 5 for the case of Exp. 1a (clean speech, random FER). The modified appendix is called G.722 App. IV+. It improves slightly the quality of G.722 Appendix IV as expected.

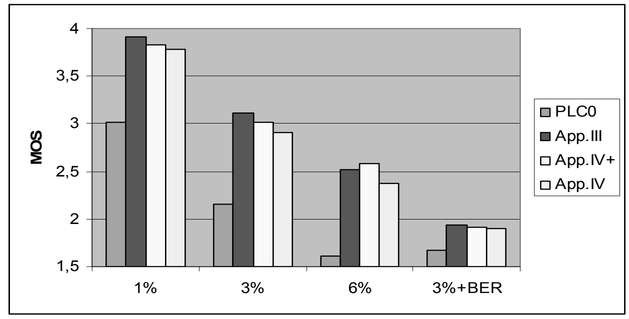

Figure 5: Additional test result (Exp1a including G.722 App. IV+).

Similar improvements were observed in all other experiments.

6. CONCLUSION

This article presented G.722 Appendix IV, which is a low-complexity PLC algorithm for G.722. This algorithm has been designed to produce good quality in the presence of frame erasure without increasing computational complexity and adding limited extra ROM requirements, to facilitate its integration in low capacity devices. Note that frame lengths multiple of 10 ms are supported. Subjective test results show that the quality of G.722 Appendix IV can be further improved by some minor modifications.

REFERENCES

[1] ITU-T Rec. G.722, “7 kHz audio coding within 64 kbit/s,” Nov. 1998.

[2] G.722 Appendix III “A high quality packet loss concealment algorithm for G.722,” Nov. 2006.

[3] J. Thyssen, R. Zopf, J-H Chen and N. Shetty, “A Candidate for the ITU-T G.722 Packet Loss Concealment Standard,” in Proc. *ICASSP 2007*, pp. IV-549, April 2006.

[4] G.722 Appendix IV “A low-complexity algorithm for packet loss concealment with G.722,” Nov. 2006, http;//www.itu.int/md/dologin_md.asp?id=T05-SG16-061114-TD-PLEN-0356!A3!MSW-E&type=mitems

[5] “Term of Reference (ToR) and time schedule for the G.722 Packet Loss Concealment (G.722 PLC) standardization,” Doc. TD235R1 (PLEN/16), ITU-T SG16 Rapporteur meeting Geneva, June 2006.

[6] “Annex Q10.H Overview of G.722 PLC selection phase,” in document TD 335 (PLEN/16), ITU-T SG16 plenary meeting, Geneva, Nov. 2006.

[7] “Global Analysis for the G.722 PLC Selection Test,” Doc. AH-06-38r2, ITU-T Q.7/12 Rapporteur meeting, Lannion, Oct. 2006.

[8] “Quality Assessment Test Plan for ITU-T G.722 PLC selection phase,” Doc. TD 175 (WP 3/16), ITU-T SG16 plenary meeting, Geneva, Nov. 2006.