IMPROVED PACKET LOSS CONCEALMENT FOR PCM VoIP

Q. Li, M. Elsabrouty and M. Bouchard
School of Information Technology and Engineering, University of Ottawa, 800 King Edward,
Ottawa, Ontario, K1N 6N5 {qili, melsabro, bouchard}@site.uottawa.ca

1. INTRODUCTION

Voice-over-IP (VoIP), the transmission of packetized voice over IP networks, is gaining much attention as a possible alternative to conventional public switched telephone networks (PSTN). However, impairments present on IP networks, namely jitter, delay and channel errors can lead to the loss of packets at the receiving end. This packet loss degrades the speech quality. Model-based speech coders, such as the International Telecommunication Union (ITU-T) G.729A and G.723.1 standards, have been extensively used for speech coding over IP networks because of their low bit rates requirements (5.3 to 6.4 kbit/s for G.723.1 and 8 kbit/s for G.729A). Also, they have an inherent ability to recover from erasure: using their model-based structure they include a built-in packet loss concealment scheme which makes their quality drop slowly with increasing amount of packet loss. However, their model requires a few frames to adapt to the transition from a concealed state to a correct state. Thus, model-based speech coders actually tend to corrupt a few good packets before recovery, as a result of a phenomenon known as "State Error"[1].

On the other hand, speech coded with Pulse Code Modulation (PCM, ITU-T G.711, 64 kbit/s), although having a higher quality compared to G.729A and G.723.1 in the periods of normal operation, does not have the built-in ability to conceal erasure. This results in a serious drop in speech quality during loss periods. Yet, PCM-based coders can recover from packet loss faster than model-based coders, since the first speech sample in the first good packet restores the speech to its original quality. With a suitable packet loss concealment scheme, PCM would make a very viable alternative to G.729A or G.723.1 for VoIP, considering its low complexity, its superior quality under normal conditions and its good performance in tandem coding. This paper introduces a high-performance concealment algorithm for PCM-coded speech.

2. A NEW PCM PACKET LOSS CONCEALMENT ALGORITHM

The new linear prediction based concealment technique is using a linear prediction with a fairly large order filter to model the speech:

$$S(n) = \sum_{i=1}^{P} \left(a_i \times S(n-i)\right) + b(n) \qquad (1),$$

where $S(n)$ is the n^{th} speech sample, P is the prediction order (set to 50), a_i are the linear prediction coefficients, and $b(n)$ is the residual signal of the prediction. As can be seen from (1), the current speech sample $S(n)$ is composed of two components. The first component is the predictable part carrying the information of the vocal tract along with the correlation between the current sample and the 50 previous ones. The second component is the residual signal $b(n)$ that contains the current unpredictable excitation.

In the case of a lost packet, the previous speech samples are available and thus the predictable term in (1) can be computed. However, the residual signal $b(n)$ is unknown to the receiver side. In this case, a good choice can be to use a small percentage of a pitch-predicted signal as $b(n)$, or in other words as the input excitation for the synthesis system. Here, the pitch-predicted signal refers to a Reverse-Order Pitch Period Replication (RORPP) of the lost frame, estimated in a manner similar to the concealment algorithm implemented in the ITU-T G.711-Annex A [2]. Thus, using a small percentage of the pitch-predicted signal, equation (1) can be re-written as:

$$S(n) = \sum_{i=1}^{P} \left(a_i \times S(n-i)\right) + \hat{S}(n) \times G \qquad (2),$$

where $S(n)$ denotes the concealed speech sample and $\hat{S}(n)$ is the pitch-predicted signal obtained from the ITU-T G.711-A (RORPP) concealment scheme. A value of G = 0.01 was found to give the best results in practice [3].

In "best effort" IP networks, future packets often arrive early. Thus it becomes possible to use future packets from the jitter buffer to predict a lost packet. A Hamming window can be applied to combine the prediction from past packets and the prediction from a future packet (or more, if available). Also, to provide a better approximation of the original signal, the concealment algorithm can be modified to perform a weighted summation of the speech predicted by linear prediction (i.e. equation (2) for the case where future samples are not considered) and the pitch-based prediction, using a Voiced/Unvoiced (V/UV) speech classification to tune the parameters. For our experiments, for the V/UV classification a simple scheme based on the linear prediction residual energy was used. A threshold was determined, based on the fact that voiced speech tends to be more energetic in the residual signal. The proposed packet loss concealment algorithm for PCM speech thus becomes:

$$S^p(n) = \sum_{i=1}^{P} \left(a_i \times S^p(n-i)\right) + \hat{S}(n) \times G \quad (3)$$

$$S^f(n) = \sum_{i=1}^{P} \left(a_i \times S^f(n+i)\right) + \hat{S}(n) \times G \quad (4)$$

$$S_v(n) = \begin{pmatrix} \left(\alpha_v \times S^p(n) + \beta_v \times \hat{S}(n)\right) \times \text{Hamming win. } 2^{\text{nd}} \text{ half}) \\ + \left(\alpha_v \times S^f(n) + \beta_v \times \hat{S}(n)\right) \times \text{Hamming win. } 1^{\text{st}} \text{ half}) \end{pmatrix} \quad (5)$$

$$S_{uv}(n) = \begin{pmatrix} \left(\alpha_{uv} \times S^p(n) + \beta_{uv} \times \hat{S}(n)\right) \times \text{Hamm. win. } 2^{\text{nd}} \text{ half} \\ + \left(\alpha_{uv} \times S^f(n) + \beta_{uv} \times \hat{S}(n)\right) \times \text{Hamm. win. } 1^{\text{st}} \text{ half} \end{pmatrix} \quad (6),$$

where $S_v(n)$ and $S_{uv}(n)$ are the final concealed signals for voiced and unvoiced frames, respectively, $S^p(n)$ and $S^f(n)$ are the linear prediction result from the past and future samples, respectively, α_v and β_v are summation weights for voiced frames, and α_{uv} and β_{uv} are summation weights for unvoiced frames. The best results were obtained with $\alpha_v = 0.9$ and $\beta_v = 0.1$, $\alpha_{uv} = 0.6$ and $\beta_{uv} = 0.4$.

3. PERFORMANCE OF THE PROPOSED ALGORITHM

The new algorithm was compared to the ITU-T G.711-Appendix A concealment tool and to the packet repetition method. The test was performed on a set of speech files from four speakers (two males and two females) referred to in the results as M1, M2, F1 and F2. For each of those speakers, 10 speech files were used, each containing two sentences in English for a duration of 8 sec. Frames of 80 samples (10 ms) were used. The format of the files was linear PCM, with 8 kHz sampling rate. The files were taken from the ITU-T supplement P.23. The assessment tool used to evaluate the results of the concealment techniques was the Perceptual Estimation of Speech Quality (PESQ) standard P.862 developed by the ITU-T [4]. This tool has shown to give reliable estimation of subjective quality tests. The scores produced by the PESQ are in the range -0.5 (very poor quality) to 4.5 (very good quality), similar to the standard Mean Opinion Score (MOS) scale.

Random loss pattern tests were performed for loss rates of 5%, 10 % and 25%. Table 1 summarizes the average PESQ results for those three loss rates. It can be seen from the above figures that the performance of the new algorithm is superior to both the existing ITU-T G.711A method and the packet repetition method. A significant and almost steady margin appears as a difference between the new proposed algorithm and the ITU-T G.711A method, and both methods perform much better than the packet repetition technique. The margin between the proposed method and the ITU-T G.711A represents the performance gain of incorporating the linear prediction model (and the possible use of future samples) with the pitch-repetition based concealment.

4. CONCLUSION

In this paper, a new concealment algorithm for PCM packetized speech was presented. The model provides very encouraging results for the idea of combining a pitch prediction along with a high-order linear prediction to produce the concealed speech samples. Future samples are taken into account (optionally), as well as the adaptation of the weighting coefficients α and β based on a V/UV classification. The PESQ-MOS scores obtained for the random loss tests have shown that the proposed algorithm exhibits a superior high-quality concealment performance in all cases, when compared to an existing commercial method (packet repetition) or to the ITU-T G.711 A concealment technique.

REFERENCES

[1] ITU-T G.711 Appendix A (2000). A high quality low-complexity algorithm for packet loss concealment with G.711. ITU-T Recommendation.
[2] Montminy, C. and Aboulnasr, T. (2000). Improving the performance of ITU-T G.729A for VoIP. International Conference on Multimedia Exposition 2000 (ICME 2000), vol. 1, 433–436.
[3] Elsabrouty, M., Bouchard, M. and Aboulnasr, T. (2004). Receiver-based packet loss concealment for pulse code modulation (PCM G.711) coder. Signal Processing, vol. 84, 663-667.
[4] ITU-T P.862 (2000). Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone network and speech codecs. ITU-T Recommendation.

Table 1 Average PESQ Results

File	Test	New Algorithm	G.711A	Packet Repetition
M1	5% loss rate	3.98	3.45	3.05
	10% loss rate	3.61	3.09	2.62
	25% loss rate	3.20	2.60	2.28
M2	5% loss rate	3.80	3.41	2.81
	10% loss rate	3.55	3.12	2.61
	25% loss rate	3.02	2.63	2.24
F1	5% loss rate	3.83	3.36	2.88
	10% loss rate	3.53	2.93	2.50
	25% loss rate	3.05	2.58	2.15
F2	5% loss rate	3.76	3.31	2.86
	10% loss rate	3.53	2.87	2.36
	25% loss rate	2.97	2.43	1.83