ITG-Fachtagung
Sprachkommunikation 2006

Packet Loss Concealment with Side Information for Voice over IP in Cellular Networks

Frank Mertz and Peter Vary
Institute of Communication Systems and Data Processing (IND)
RWTH Aachen University, D-52056 Aachen, Germany
{mertz, vary}@ind.rwth-aachen.de

Abstract

In this paper we present a concept for robust speech transmission on packet based cellular networks with packet losses. We transmit selected side information of low bit rate to assist the receiver's packet loss concealment routine in reconstructing lost frames of standard ACELP based speech codecs, e.g., AMR Wideband. The side information indicates which particular estimation technique the receiver should use to restore speech parameters of the respective frame. Further improvement can be achieved by additionally transmitting the coarsely quantized estimation error. The achievable quality gains are evaluated by parameter SNR, spectral distortion, and Wideband PESQ measures.

1. Introduction

Voice over IP (VoIP), i.e., packet-switched speech transmission, has become increasingly important over the last years. Currently deployed 3G mobile communication systems (e.g. UMTS) even facilitate the introduction of Voice over IP services over mobile radio channels. However, the problem in Voice over IP transmission over heterogeneous networks are packet losses caused by network congestion or bit errors on a wireless link. VoIP applications in cellular networks utilize standardized codecs that originally have been designed for circuit-switched wireless networks with limited bit rates, e.g., ITU-T G.729A or the Adaptive Multi-Rate Wideband (AMR-WB) codec. These codecs are sensitive to packet loss, because the inherent ACELP codec structure leads to error propagation, particularly through an incorrect update of the adaptive codebook. On the other hand, codecs specifically designed for packet switched networks, e.g., the Internet Low Bit Rate Codec (iLBC, IETF RFC 3951), achieve a higher robustness against packet loss by avoiding dependencies between adjacent encoded speech frames. However, this advantage is gained at the expense of a higher data rate. In this paper we use additional bit rate to enhance the robustness of existing codecs towards packet loss by transmitting some selected side information to assist the receiver's packet loss concealment.

There are different strategies to combat packet loss in VoIP applications. Sender-driven approaches are able to limit the impact of packet loss to some extent, e.g., by transmitting a copy of the current speech frame in a following packet or by applying other channel coding schemes. However, the benefit of these methods is gained at the expense of a considerable increase in bit rate which may not be tolerable for wireless transmission channels. Methods have also been proposed which only transmit partial information on some speech parameters in following packets to limit the overall bit rate, e.g. [1],[2]; the missing information will be estimated by the receiver's *packet loss concealment* routine. Finally, solely receiver-based concealment approaches utilize received preceding and succeeding frames to reconstruct a lost frame without side information, as, e.g., shown in [1],[3],[4],[5]. For cellular networks with limited bit rates, a solution for robust packet based speech transmission lies in between the bit rate intensive sender-driven and solely receiver-based approaches. In the fol-

lowing, a combined concept will be presented, which will be referred to as *sender-assisted packet loss concealment (PLC)*.

2. Sender-assisted PLC Approach

The concept of sender-assisted packet loss concealment is to transmit selected side information in succeeding packets to assist the receiver's concealment routine in estimating the lost speech parameters. Two *types of side information* are considered:

1) side information on which estimation technique to use for error concealment in the receiver (e.g., extra-, interpolation),
2) side information to improve the estimation (e.g., quantized estimation error).

The appropriate estimation technique for a parameter depends on the current signal structure. While the choice can be based on the voicing state of adjacent frames [3], some explicit side information on which estimation technique to use for a specific speech frame can further improve the estimation. We compared different estimation techniques for each parameter of the AMR-WB codec and identified which techniques to implement in the receiver. From this set, the optimal technique for individual frames is determined at the sender and transmitted as side information (type 1). Knowing which estimation technique the receiver will use in case of loss, the coarsely quantized estimation error can be optionally transmitted as additional information to improve the concealment (type 2). These types of side information require a lower additional bit rate than approaches that transmit copies of speech parameters.

2.1 Simulation Settings

For the simulations presented in the following sections, the test files of the TIMIT database have been used (sampling frequency: 16 kHz). For each speaker, several short files were combined in a single file, resulting in 168 files, each of about 8 to 12 sec length. The files were encoded by the Adaptive Multi-Rate Wideband codec with 23.05 kbit/s. The performance of the different concealment methods will be assessed by the parameter SNR (pSNR) or the mean spectral distortion ($\overline{\mathrm{SD}}$). Furthermore, we will discuss the impact on the resulting speech quality by presenting Wideband PESQ [6] measurements for 10% single frame losses in the speech files. The results are shown in Fig. 1 and will be discussed in the following sections.

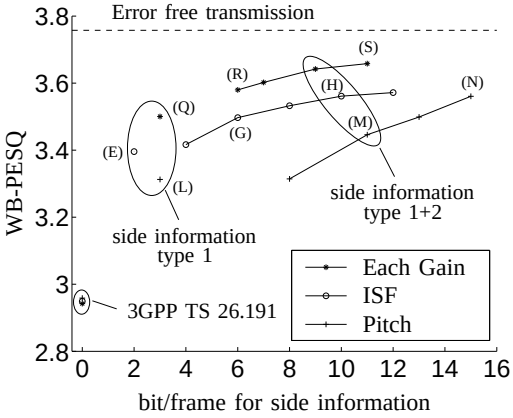

Fig. 1: WB-PESQ measurements for different sender-assisted concealment approaches: Each AMR-WB parameter considered separately, respective other parameters received correctly.

3. Concealment Methods

In the following, the different parameters of the AMR-WB speech codec are considered separately. In these studies, the respective other parameters have been assumed as received correctly in order to focus only on the influence of the considered parameter.

3.1 Spectral Envelope – ISF

The AMR-WB codec uses *immittance spectral frequencies* (ISF) as representation for the LP coefficients to describe the spectral envelope of a speech frame. For transmission, a residual $\mathbf{r}(n)$ is calculated recursively from the mean removed ISF vector $\mathbf{q}(n)$ according to:

$$\mathbf{r}(n) = \mathbf{q}(n) - \overline{\mathbf{q}} - \frac{1}{3} \cdot \mathbf{r}(n-1) \qquad (1)$$

with $\overline{\mathbf{q}}$ a constant mean (expectation) ISF vector. Therefore, a frame loss always leads to an error propagation of one further frame. Table I shows the quality that can be achieved by various concealment approaches (A-F). In the standard concealment approach (A) [7], the lost ISF vectors are estimated by shifting the

past ISFs towards their mean. However, if the frame following a loss is available, it can be utilized for the concealment. Assuming a loss of a single frame n, we propose the following interpolation function for the lost ISF vector:

$$\hat{\mathbf{q}}(n) = \alpha \cdot \mathbf{q}(n-1) + (1-\alpha) \cdot \mathbf{q}(n+1) \quad (2)$$

with parameter $\alpha \in [0,1]$ determining a weighting between the previous and the following value. A linear interpolation of the ISF vectors is achieved for $\alpha = 0.5$ (C). Note that the following ISF vector $\mathbf{q}(n+1)$ is not explicitly known at the receiver, only the received residual $\mathbf{r}(n+1)$. In [4] it was proposed to first extrapolate the ISF as in the standard approach, then calculate $\mathbf{q}(n+1)$, and finally linearly interpolate $\mathbf{q}(n-1)$ and $\mathbf{q}(n+1)$ (B). However, assuming the interpolation with parameter α resembles the actual ISF vectors close enough, we obtain a closed mathematical solution to Eq. (2) by substituting $\mathbf{q}(n-1)$ and $\mathbf{q}(n+1)$ according to Eq. (1), which finally yields

$$\hat{\mathbf{q}}(n) = \overline{\mathbf{q}} + a \cdot \mathbf{r}(n-2) + b \cdot \mathbf{r}(n-1) + c \cdot \mathbf{r}(n+1)\,;$$
$$a = \frac{\alpha}{\alpha+2};\; b = \frac{10\alpha - 1}{3\alpha + 6};\; c = \frac{3-3\alpha}{2+\alpha}. \quad (3)$$

For the proposed sender-assisted (SA) approach of packet loss concealment, simulations with different sets of α values have been carried out (D, E, F). For each frame, the optimal α has been determined as the one that minimizes the spectral distortion between the spectra belonging to $\hat{\mathbf{q}}(n)$ and $\mathbf{q}(n)$. A choice of 4 values for α proves reasonable (E), i.e., 2 additional bits per frame have to be transmitted. The quality can be further improved by transmitting the quantized estimation error vector $\mathbf{e}_q(n) = \mathbf{q}(n) - \hat{\mathbf{q}}(n)$. The results are depicted in Fig. 1 for several bit rates and show a further noticeable improvement for, e.g., 4 additional bits/frame, i.e., a total side information for the ISF of 6 bit/frame (G).

3.2 Pitch Lag

The parameter SNR and WB-PESQ values resulting from different approaches (I-L) are listed in Table II. In the standard concealment method (I) [7], the pitch lags of a lost speech frame (4 sub-frames) are either repeated or randomly varied, depending on the voicing state. For the sender-assisted (SA) approach, we considered the following estimation techniques for lost pitch lags: standard concealment according to [7] (I), linear interpolation (J), and a replacement approach, where the first N sub-frames will be estimated by the preceding value and the remaining sub-frames by the first pitch value of the following frame. The choice of $N \in [0,4]$ is included in the side information. The estimation technique that produces the smallest mean-square error for the pitch lags of a frame is chosen and its index is transmitted to the receiver in a succeeding packet (K, L). Method L requires 3 additional bits per frame and leads to a considerable improvement over [7]. Further improvement can be achieved by transmitting the quantized estimation error with 8 bit/frame (see Fig. 1, M).

TABLE I
ISF ESTIMATION: PERFORMANCE OF DIFFERENT CONCEALMENT APPROACHES

	Concealment Method	$\overline{\text{SD}}$ [dB]	WB-PESQ
A	3GPP TS 26.191 [7]	4.22	2.94
B	Extra-/Interpolation [4]	3.54	3.24
C	Linear Interpolation: $\alpha = 0.5$	3.61	3.26
D	SA, 2 values of α: 0.5;0.8	3.22	3.33
E	SA, 4 values of α: 0.3;0.5;0.7;0.9	3.02	3.40
F	SA, 8 values of α: 0.3;0.4;...;1.0	2.99	3.40

TABLE II
PITCH ESTIMATION: PERFORMANCE OF DIFFERENT CONCEALMENT APPROACHES

	Concealment Method	pSNR [dB]	WB-PESQ
I	3GPP TS 26.191 [7]	6.45	2.96
J	Linear Interpolation	7.01	3.06
K	SA, 2 techniques (1 bit): I, J	8.50	3.13
L	SA, 7 techniques (3 bit): I, J, replacement ($N \in [0,4]$)	9.96	3.31

3.3 Adaptive and Fixed Codebook Gains

The achievable quality of different methods (O-Q) is shown in Table III. In the standard concealment method (O), the gains of adaptive and fixed codebook are estimated by attenuated values from the previous sub-frames. While signal muting is necessary in cases of several consecutive frame losses, for short losses of

TABLE III
GAIN ESTIMATION: PERFORMANCE OF DIFFERENT CONCEALMENT APPROACHES

Concealment Method		pSNR [dB]		WB-PESQ
		g_a	γ	
O	3GPP TS 26.191 [7]	N/A	N/A	2.94
P	SA, 11 techniques (4 bit)	11.41	4.69	3.51
Q	SA, 8 techniques (3 bit)	11.19	4.61	3.50

1-2 frames this attenuation leads to noticeable and unnecessary amplitude fluctuations in voiced speech segments, as shown in [3].
For the sender-assisted (SA) concealment approach (P), we use the following estimation techniques: interpolation, sub-frame replacement by previous or following gains (as for pitch lags) with $N_1 \in [0, 4]$, or replacement by the mean gains of the previous or following frame with $N_2 \in [0, 4]$. The decision for an estimation technique is made separately for the adaptive codebook gain g_a and the received correction factor $\gamma = g_c/g_c'$ between the fixed codebook gain g_c and its prediction g_c'. Even with a restriction of N_1 to 2 values $\{0, 4\}$ (Q) already a considerable improvement over [7] (O) is achieved. Further improvement requires at least 3 more bits/frame for each gain to transmit the quantized estimation error (Fig. 1, R).

4. Speech Quality Measurements

We finally combined the approaches for the different codec parameters of Fig. 1 to evaluate the overall quality improvement by the proposed sender-assisted packet loss concealment for the example of 10% single frame losses. Table IV shows the Wideband PESQ [6] results for different bit rates of additional side information. In comparison to the standard approach [7], a considerable quality improvement is already gained by transmitting which estimation techniques to use for concealment (11 bit/frame), which can be further increased by transmitting quantized estimation errors. The fixed codebook (FCB) excitation either has been estimated by a random sequence [7] or, for comparison, assumed as correctly received. The transmission of side information for this parameter will be subject to further studies.

TABLE IV
QUALITY OF COMBINED APPROACH

Concealment Method (see Tab. I-III, Fig. 1)	WB-PESQ	
	correct FCB	random FCB
3GPP TS 26.191 [7]	2.240	2.175
SA, 11 bit/frame (E, L, Q)	2.857	2.529
SA, 29 bit/frame (G, M, R)	3.040	2.609
SA, 47 bit/frame (H, N, S)	3.242	2.675

5. Conclusion

We presented a new concept of sender-assisted packet loss concealment which is based on the transmission of side information to improve the concealment of lost frames at the receiver. Two types of side information have been considered, first information on what estimation technique is optimal for each codec parameter of a specific frame, and second a coarse quantization of the respective estimation error. We have shown that with an additional bit rate of only 11-29 bit per frame, a considerable quality improvement can be gained for single frame losses. Further studies will be made to extend these approaches to longer frame loss lengths.

References

[1] I. Johansson, T. Frankkila, and P. Synnergren, "Bandwidth efficient AMR operation for VOIP," in *IEEE Workshop on Speech Coding*, Tsukuba, Ibaraki, Japan, Oct. 2002.

[2] L. Tosun and P. Kabal, "Dynamically Adding Redundancy for Improved Error Concealment in Packet Voice Coding," in *EUSIPCO 2005*, Antalya, Turkey, Sep. 4-8 2005.

[3] F. Mertz, H. Taddei, I. Varga, and P. Vary, "Voicing Controlled Frame Loss Concealment for Adaptive Multi-Rate (AMR) Speech Frames in Voice-over-IP," in *Eurospeech 2003*, Geneva,Switzerland, 2003.

[4] T. Fingscheidt and J. G. Perez, "An Interpolative Decoding Approach for Speech Streaming Services and Voice Over IP," in *Internat. ITG Conference on Source and Channel Coding*. Berlin, Jan. 2002.

[5] J. Wang and J. Gibson, "Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks," in *ICASSP 2001*, vol. 2, Salt Lake City, Utah, USA, 7-11 May 2001.

[6] ITU-T Rec. P.862.2, "Wideband extension to Rec. P.862 for the assessment of wideband telephone networks and speech codecs (prepublished)," 2005.

[7] 3GPP, "TS 26.191: Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Error concealment of erroneous or lost frames."