BACKWARDS-COMPATIBLE ERROR PROPAGATION RECOVERY FOR THE AMR CODEC OVER ERASURE CHANNELS

Angel M. Gomez, José L. Pérez-Córdoba

Dpt. of Signal Theory, Telematics and Communications, and CITIC-UGR
University of Granada, Spain

Bernd Geiser

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany

ABSTRACT

This paper presents a recovery scheme for the error-propagation distortion which frequently appears after a frame erasure in CELP-based speech coders, in particular the AMR codec. The extensive use of predictive filters and parameter encoding allow a high-quality speech synthesis in these codecs, but makes them more vulnerable to frame erasures. Thus, when a frame is lost, an additional distortion appears in the subsequent frame, although that was correctly received, further degrading the speech quality. This degradation can also propagate over several frames, being even more damaging than the loss itself. This well known fact has motivated the development of techniques which prevent or mitigate the error propagation. Nevertheless, the previously proposed methods in some respect modify the transmission scheme (by including additional frames, FEC codes, etc.) making them incompatible with the original decoder. In this work, we apply a steganographic technique to embed recovery data to assist the decoder after a frame loss. This data mainly consist of resynchronization pulses and correction vectors for the excitation signal and the spectral envelope, respectively. PESQ results confirm that our proposal achieves a higher robustness against error propagation while the full backwards-compatibility with the AMR standard is retained.

***Index Terms*—** speech coding, ACELP, frame erasure, error propagation, pulse codebook, data hiding, steganography

1. INTRODUCTION

The Adaptive Multi-Rate (AMR) codec is a widely used narrowband codec which provides a high-quality speech synthesis at a remarkably low bitrate. Nevertheless, due to the extensive use of predictive filters, this codec is relatively vulnerable to frame erasures, making it unsuitable for packet-based transmission as involved in the thriving new VoIP services. One of these filters, the long-term prediction (LTP) filter, which is used to build up the adaptive codebook (ACB), can exceed the boundaries of signal frames and it is considered the main source of error propagation. This kind of distortion appears after a frame erasure, when the LTP filter requires samples from a frame that has not been received. Since these samples have been replaced by a concealment algorithm, the ACB codebook in the decoder desynchronizes from the encoder, causing a degradation which can propagate over several consecutive frames [1].

Traditionally, research efforts have focused on concealing the erased frames themselves but, in recent years, subsequent error propagation has also been noted as a remarkable source of degradation and a number of authors has proposed different techniques to counteract it. The most straightforward approach simply consists of avoiding inter-frame predictions (such as LTP) as in the Internet low bitrate codec (iLBC) [2] but that implies a considerable increase in bit rate [3]. This can be alleviated to some extend by a hybrid solution [4] which combines iLBC (independent) frames and CELP (dependent) frames. An alternative approach is sending some additional recovery information. This redundancy data can be added at the packetization level, as in [5] or by repeating all, or a part, of the encoded parameters [3, 6]. In [7] additional frames are used to re-initialize the decoder in case of packet loss but, since this strategy implies doubling the bit rate, the redundancy information is only sent for onset frames, as they are considered especially vulnerable to previous losses [8]. On the other hand, the solutions proposed in [9, 10, 11] analyze the speech signal in order to extract some FEC parameters about glottal pulse structure, which can be encoded using fewer bits than a regular frame. Similarly, in a previous work [12], we proposed a novel technique for error propagation mitigation which consists of the computation and transmission of a pulse-based representation of the previous frame excitation. This alternative representation allows the decoder to resynchronize its internal memory and to obtain a similar performance as with a fully synchronized ACB [12]. This approach, however, as all aforementioned techniques, also causes an inherent, yet smaller, increase of the bit rate.

In order to avoid this overhead, we propose in this work the use of a data hiding (or steganographic) technique which allows us to embed the recovery information into the codec bitstream itself. CELP bitstream data hiding principles were proposed by Lu et al. [13] where a rather low capacity of 37 hidden bit/s was achieved. This capacity was further extended by the steganographic ACELP codec proposed in [14] allowing bit rates of several hundred of bps without compromising the quality of the coded speech signal. Here, we will exploit an advanced data hiding technique for ACELP codecs [15] that embeds bit rates from 200 bps up to 2 kbps.

In this work, we will focus on the AMR codec, in particular on its highest quality mode (12.2 kbps). However, the scheme proposed could be extended to other ACELP-based coders. As we will show, by embedding hidden recovery data to assist the decoder after a frame loss, we can obtain an AMR codec that is robust against propagation errors and, at the same time, maintains full compatibility with the AMR standard.

The remainder of this paper is organized as follows. Section 2 describes the recovery data to assist the decoder against the error propagation after a frame loss, while their encoding and transmission is outlined in Section 3. Section 4 is devoted to the experimental framework and results. Finally, conclusions are drawn in Section 5.

This paper has been supported by the Spanish Ministry of Education through project MICINN TEC2010-18009.

2. RECOVERY INFORMATION AFTER A PACKET LOSS

The AMR codec is based on the code-excited linear prediction (CELP) paradigm, where speech is obtained by filtering an excitation signal, $e(n)$, through a short-term linear prediction (LP) filter, $1/A(z)$. This excitation $e(n)$ is the result of a long-term prediction (LTP) filter applied over a code vector $e_c(n)$, also termed residual signal,

$$\begin{aligned} e(n) &= g_a \sum_{k=-q}^{q} p_k e(n-(T+k)) + g_c e_c(n) \\ &= g_a e_a(n) + g_c e_c(n), \end{aligned} \tag{1}$$

where T, p_k, g_a and g_c are the parameters of the LTP filter, namely lag-delay, long-term coefficients, and adaptive and code gains, respectively. Alternatively, the excitation signal can be seen as a weighted sum of two different components: $e_a(n)$, an adaptive vector that models the long-term correlation in excitation, and $e_c(n)$ which is chosen from a fixed codebook and represents the remaining residual signal.

Although encoding is carried out on a frame or sub-frame basis, there exist inter-frame dependencies during the decoding process, as the excitation signal of a frame is later used by the LTP filter to compute the adaptive vector of the next frame. These inter-frame dependencies endanger the codec performance in packet-based networks. When a frame is lost, a concealment algorithm replaces it in order to minimize the degradation on the perceptual quality. Since the concealed signal is not identical to the transmitted one, the decoder desynchronizes from the encoder and a distortion appears which can propagate over several subsequent correctly received frames.

In addition, LPC coefficients ($A(z)$ filter) are also predictively encoded in AMR. Fortunately, this encoding process is performed in such a way that the effects of a frame erasure are confined to one frame after the loss. Concretely, the coefficients a_k are transformed into linear spectral frequencies (LSF), f_i, which are quantized after a 1st order MA prediction:

$$r_i(m) = f_i(m) - \bar{f}_i - 0.65\hat{r}_i(m-1) \qquad i = 1, \ldots, P, \tag{2}$$

where P is the number of LSF coefficients per frame ($P = 10$ in AMR), m is the frame index, $r_i(m)$ is the resulting prediction residual, $\bar{f}_i$ is the mean LSF value and $\hat{r}_i(m-1)$ is the quantized residual at the past frame[1]. As can be observed, after a frame loss, the previous residual $\hat{r}_i(m-1)$ is not available (only its concealed version is) causing a degradation on the spectral envelope which combines with that in the excitation signal.

In order to alleviate these problems we propose the transmission of selected side data to assist the decoder after a frame erasure. The data includes information about the spectral envelope and the excitation signal.

2.1. Spectral envelope

As we mentioned before, LSF decoding only depends on the previous frame residual. This means degraded LSFs could be easily improved by an interpolation-based approach, as in [16], using the next frame parameters. However, this would introduce an additional delay during decoding. To avoid it, we can try to predict the value of the previous frame residual from the current frame LSF residual (or residuals in the case of the 12.2 kbps mode). This can be done since the simple MA prediction applied in AMR not only limits the effect of a frame loss over the spectral envelope, but also leaves some redundancy across the residuals. Thus, each component of the previous frame residual set can be linearly predicted from all the elements of the current frame residual as,

$$\tilde{r}_i(m-1) = \sum_{k=1}^{P} \alpha_k r_k(m) + \beta \qquad i = 1, \ldots, P \tag{3}$$

where the prediction coefficients, α_k and β, are obtained by means of a linear regression applied over a training database. Of course, some error δ_i between the predicted residual $\tilde{r}_i$ and the actual previous one, $\hat{r}_i$, is expected. However, preliminary tests showed that the variance of this error is significantly lower that the variance of the residual itself. Hence, in order to improve the spectral envelope after a frame erasure, this error could be transmitted along with the current frame using a few bits.

2.2. Excitation signal

In a previous work [12], we proposed to send a multipulse encoded version of the previous frame excitation as a way to resynchronize the LTP filter. This alternative excitation is composed of several pulses and is intended to be used as a replacement of that provided by the standard PLC algorithm.

Under this approach, previous frame excitation samples are seen as a memory where some pulses can be set. These will be later transformed by the lag filter, scaled by the adaptive gain and added to the code vector (correspondingly scaled by the respective gain) generating the excitation of the current frame. Pulse parameters are thus optimized in order to provide the minimum error between the resulting synthesized signal and the original one. To this end, the least square error (LSE) criterion is applied,

$$\varepsilon = \sum_{n=0}^{N-1} \big(s(n) - h(n) * \hat{e}(n)\big)^2 \tag{4}$$

where $h(n)$ is the impulse response of the LP filter, $s(n)$ is the target signal and N is the number of samples per frame. Provided the perceptual filtering and code vector contributions are previously removed from the target signal ($s_{zi}(n)$) as in [12], the excitation signal, $\hat{e}(n)$, can be simplified as,

$$\hat{e}(n) = \sum_{l=0}^{L-1} b_l h_{LT}(n, p_l) \tag{5}$$

where $h_{LT}(n, p_l)$ is the response of the time-variant LTP filter[2] to a unitary pulse at position p_l. The square error between the synthesized signal and the target is therefore given by,

$$\begin{gathered} \varepsilon = \sum_{n=0}^{N-1} \Big(s_{zi}(n) - \sum_{l=0}^{L-1} b_l \cdot g_{p_l}(n)\Big)^2 \\ \text{with} \qquad g_{p_l}(n) = h(n) * h_{LT}(n, p_l). \end{gathered} \tag{6}$$

Thus, provided a set of L pulse positions p_k ($k = 0, \ldots, L-1$) is known, optimal amplitudes b_k are obtained by means of the following set of equations:

$$\sum_{k=0}^{L-1} b_k^* \phi_{p_k p_j} = c_{p_j}, \qquad 0 \le j \le L-1 \tag{7}$$

[1] In the particular case of AMR 12.2 kbps, two LSF sets are obtained per frame. Both of them are quantized by applying eq. (2) but always using the second set's residual as $\hat{r}_i(m-1)$.

[2] Different LTP coefficients are applied in each subframe, resulting in a time-variant linear filter for the complete frame.

which is identical to that from the multipulse encoding except the LP impulse response is now convolved with the LTP one,

$$\phi_{p_k p_j} = \mathbf{\Phi}[k,j] = \sum_{n=0}^{N-1} g_{p_k}(n) g_{p_j}(n),$$

$$c_{p_j} = c[j] = \sum_{n=0}^{N-1} s(n) g_{p_j}(n). \tag{8}$$

Although a faster resynchronization is obtained as the number of pulses increases [12], in this work we will only consider a single pulse. This significantly reduces the complexity of the scheme and enables us to obtain optimum pulse amplitudes and positions (in contrast to the multipulse case where a suboptimal algorithm is used [17]). Provided that only one pulse is needed, the set of (7) reduces to,

$$b^* \cdot \phi_{pp} = c_p \tag{9}$$

where, for simplicity, $p = p_0$, $b = b_0$. Then, it can be proven that the optimal pulse position i^* and amplitude b_i^* are obtained as,

$$i^* = \underset{i}{\operatorname{argmax}}(c_i^2/\phi_{ii}), \qquad b_i^* = c_{i^*}/\phi_{i^* i^*} \tag{10}$$

Nevertheless, as we will show in the following section, these equations are only useful to compute the codebook which we will use during quantization. LTP pulse compensation is finally achieved by a simple search of the best pulse (in an LSE sense) in this pulse codebook.

3. CODING AND TRANSMISSION

The aforementioned recovery data must be encoded and sent in every frame so that, in case of a frame erasure, these can assist the frame spectral envelope decoding and excitation resynchronization. Classical media-specific FEC codes can be used to this end. However, steganographic techniques could be used instead, not only avoiding the inherent increase in bit rate but also allowing compatibility with the standard AMR bitstream.

3.1. Quantization

The previously described LSF residual errors $(\delta_1, \ldots, \delta_P) = \boldsymbol{\delta}$ are quantized using vector quantization (VQ). A weighted distortion measure, which is identical to that described in the AMR standard for LPC quantization [18], is applied during the process. LSF residual-error codebook is obtained through the k-means algorithm applied over a training database.

A different approach is followed for pulse representation since a direct quantization could lead to severe quantization errors. Instead, a cost function for setting a pulse at an arbitrary position i with an arbitrary amplitude v when recovering the training frame m is derived as [17],

$$f_{i,v}(m) = E_{zi}(m) + v \cdot \phi_{ii}(m)[v - 2b_i^*(m)] \tag{11}$$

where $E_{zi}(m)$ is the energy of the target signal and $b_i^*(m)$ is the optimal amplitude at position i for frame m (obtained through eqn. (10)). Then, given a training database with M frames, our aim is to find a set of pulses, that is, amplitude-position pairs $J = \{\langle p,b\rangle^{(0)}, \langle p,b\rangle^{(1)}, ..., \langle p,b\rangle^{(L-1)}\}$, which minimizes the value of the total cost functional F_{pulse}, defined as,

$$F_{pulse} = \sum_{m=1}^{M} \min_{\langle p,b\rangle \in J} (f_{p,b}(m)). \tag{12}$$

In [17] the optimal cell and optimal center criteria for the Lloyd algorithm which solves this problem are described. This solution enables us to obtain a codebook of pulses (amplitude-position pairs J) which provides the minimal synthesis error over the entire training database.

As we mentioned before, joint amplitude-position codebooks require a modification on the pulse search and encoding procedure as we now have a finite set of admissible pulses. The best amplitude-position pair $\langle i,v\rangle$ in the codebook can be found by maximizing the following function [17],

$$\hat{f}\,(i,v) = v(2c_i - v\phi_{ii}). \tag{13}$$

Finally, it is worth to mention that the complexity of the corresponding search algorithm is similar to the optimization of a non-quantized single-pulse, since the highest complexity step, the computation of c_i and ϕ_{ii}, remains the same.

3.2. Steganographic embedding

Once quantized, in order to achieve backwards compatibility with the standard AMR decoder, the recovery data shall be hidden in the codec bitstream. To this end, we use the steganographic ACELP encoding technique of [15] which is based on a modified search procedure for the algebraic (fixed) codebook. Concretely, with the 12.2 kbps mode of the AMR codec, data rates from 200 bps to 2 kbps can be hidden in the bitstream. For our purposes (see Section 4), $K = 2$ (or 4) steganographic bits are transmitted per 5-ms subframe. Hence, the steganographic message μ for each subframe is defined as a K-bit binary sequence whereby the individual bits are denoted by $(\mu)_k$ with $k \in \{0, \ldots, K-1\}$. To enable the transmission of K steganographic bits per subframe, the FCB is then partitioned into $M = 2^K$ sub-codebooks that are uniquely associated with the selected message μ. The decoder can then identify the used sub-codebook and therefore decode the hidden message. Based on the standard ACELP search method from [18], the steganographic codebook search algorithm has been derived in two steps:

1. **Codebook partitioning.** First, M disjoint sub-codebooks are established by appropriately restricting the set of admissible codevectors. In particular, a specific parity condition is imposed on certain parts of the AMR bitstream:

$$(\mu)_k = \left[\mathcal{G}\left(\left\lfloor \frac{i_k}{5} \right\rfloor \right) \oplus \mathcal{G}\left(\left\lfloor \frac{i_{k+5}}{5} \right\rfloor \right) \right] \mod 2 \tag{14}$$

for the ACELP pulse positions i_k with $k \in \{0, ..., K-1\}$, where $X \oplus Y$ is the bitwise exclusive disjunction (XOR) of two binary strings and $\mathcal{G}$ represents the standardized Gray encoding of the ACELP pulse position codewords. Solving the above bitstream parity condition for the position i_{k+5} of the *second* pulse in ACELP track k, the admissible indices (and thus the possible positions) for this pulse can be computed:

$$\left\lfloor \frac{i_{k+5}}{5} \right\rfloor = \mathcal{G}^{-1}(\mathcal{G}\left(\left\lfloor \frac{i_k}{5} \right\rfloor \right) \oplus (\mu)_k + 2 \cdot j) \tag{15}$$

with $j \in \{0, \ldots, 3\}$. In fact, only the first K (out of five) pulse tracks (for the *second* pulse in that track) are restricted according to (15) to have four (out of eight) admissible pulse positions. The remaining $5 - K$ pulse tracks are not restricted here, i.e., all 8 possible pulse positions are allowed.

	bit rate	Frame erasure ratio								
		clean	*4%*	*7%*	*10%*	*13%*	*16%*	*18%*	*21%*	*23%*
AMR Standard	12.2	4.01	3.35	3.05	2.82	2.63	2.47	2.37	2.23	2.14
AMR Stego 400 bps	12.2	4.01	3.34	3.03	2.79	2.60	2.43	2.33	2.19	2.10
AMR Stego 800 bps	12.2	3.99	3.33	3.03	2.79	2.60	2.43	2.33	2.19	2.10
AMR 400 bps FEC	12.6	4.01	3.44	3.18	2.99	2.83	2.69	2.61	2.50	2.42
AMR 800 bps FEC	13.0	4.01	3.47	3.22	3.03	2.88	2.74	2.66	2.54	2.47
AMR Propos. A	12.2	4.01	3.43	3.17	2.97	2.81	2.68	2.59	2.47	2.40
AMR Propos. B	12.2	3.99	3.46	3.20	3.02	2.86	2.72	2.63	2.52	2.44

Table 1. Average PESQ scores obtained by the AMR 12.2 kbps codec (*AMR Standard*), the steganographic versions (*AMR Stego 400 bps* and *AMR Stego 800 bps*), the FEC-extended scheme with recovery information (*AMR 400 bps FEC* and *AMR 800 bps FEC*) and the combined proposals (*AMR Propos. A* and *AMR Propos. B*).

2. **Search space expansion**. Based on the chosen codebook partitioning, an FCB search strategy is devised that provides a good trade-off between speech quality and computational complexity. Thereby, the admissible values for the pulse positions i_{k+5} can be computed using eqn. (15). More details on this steganographic FCB search can be found in [15].

At the decoder, the hidden information can be retrieved directly from the AMR bitstream using eqn. (14).

4. EXPERIMENTAL RESULTS

The performance of our proposal has been tested by means of the Perceptual Evaluation of Speech Quality (PESQ) algorithm [19]. In particular, we have focused on the 12.2 kbps mode of the AMR codec, although the proposed recovery scheme can be extended to the other AMR modes as well as others ACELP-based codecs. The speech corpus is a subset of the testing part of the TIMIT database downsampled to 8 kHz with lengths artificially extended to approximately 14 seconds. This modification is performed by concatenating utterances from the same speaker, as lengths between 8 and 20 seconds are recommended for PESQ evaluation [19]. Also, utterances from female and male speakers have been balanced, providing a total of 450 test stimuli. A wide range of channel conditions has been considered, including frame erasure ratios of 4%, 7%, 10%, 13%, 16%, 18%, 21% and 23%. As in the related literature [20, 21, 3, 10, 22, 23, 24, 7], frame erasures are modeled by a random channel. Finally, in order to obtain a global score for each channel condition, utterance scores are weighted by the utterance relative length and averaged.

Table 1 shows the results obtained by the standard AMR 12.2 kbps codec (*AMR Standard*) as well as those from the techniques described in this paper. Thus, second and third rows in the table show the results obtained by the steganographic technique at hiding rates of 400 bps (*AMR Stego 400 bps*) and 800 bps (*AMR Stego 800 bps*). No recovery data is sent in these experiments, serving the provided results as a reference. The fourth and fifth rows show the results obtained using a recovery pulse of 8 bits (*AMR 400 bps FEC*)[3] and using both a pulse of 9 bits and an LSF residual error correction quantized with 7 bits (*AMR 800 bps FEC*). In both cases recovery information is sent by FEC codes in every frame, causing an increase in the final bit rate. Finally, the sixth and seventh rows show the results achieved by our proposals, where recovery data are embedded into the code vector through steganography, avoiding any increase in the bit rate. Our first proposal (*AMR Propos. A*) only uses a recovery pulse of 8 bits, requiring a hiding rate of 400 bps. The second one combines a pulse of 9 bits and an LSF residual error correction quantized with 7 bits (*AMR Propos. B*), requiring a hiding rate of 800 bps. It must be noted that, across all the bit distributions between excitation and spectral envelope recovery data, that one (9+7 bits) provided the best results.

As can be observed, steganographic techniques allow the embedding of additional data at the cost of a marginal quality reduction. This reduction is almost imperceptible in the case of clean channels. Recovery information on its part significantly improves the performance of the codec in frame loss conditions, where LTP resynchronization pulses seem to play a major role. By hiding these recovery data in the fixed code by means of steganography, a fully compatible bitstream with no out-of-band side data is obtained. In such a way, a better speech quality can be achieved by those decoders aware of the hidden data while backwards compatibility is retained with the rest. Also, it is worth to note that the proposed scheme causes no additional delay, neither for the encoder nor for the decoder.

5. CONCLUSIONS

In this paper we have presented a robust transmission scheme for AMR 12.2 which uses a steganographic technique to send recovery information which prevents error propagation after frame erasures. The recovery data included in each frame consists of a single resynchronization pulse and a residual error correction vector intended to compensate respectively the LTP filter desynchronization and the spectral envelope degradation, both appearing after a frame loss. Steganography enables the data overhead to be hidden within the algebraic code, causing no increase in bit rate and maintaining full bitstream compatibility with the standard codec. The performance of our proposal has been tested through the PESQ algorithm over a balanced subset of the TIMIT database, showing significant speech quality improvements under adverse frame-loss conditions and almost identical performance to that offered by the standard codec in clean channel conditions. This better performance has also been confirmed by preliminary results of an ongoing MUSHRA evaluation with real listeners.

[3]Note that the experimental framework used here is identical to that in [17] and results (using different numbers of bits) can be compared.

6. REFERENCES

[1] A.M. Gomez, A.M. Peinado, V. Sanchez, and A.J. Rubio, "Recognition of coded speech transmitted over wireless channels," *IEEE Trans. on Wireless Comunications*, vol. 5, pp. 2555–2562, Sep. 2006.

[2] S.V. Andersen, W.B. Kleijn, R. Hagen, J. Linden, M.N. Murthi, and J. Skoglund, "iLBC - a linear predictive coder with robustness to packet losses," *Proc. Speech Coding, 2002, IEEE Workshop Proceedings.*, pp. 23–25, October 2002.

[3] R. Lefebvre, P. Gournay, and R. Salami, "A study of design compromises for speech coders in packet networks," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. I–265–8, May 2004.

[4] J.L. Carmona, J.L. Perez-Cordoba, A.M. Peinado, A.M. Gomez, and J.A. Gonzalez, "A scalable coding scheme based on interframe dependency limitation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 4805–4808, April 2008.

[5] C.S. Xydeas and F. Zafeiropoulos, "Model-based packet loss concealment for AMR coders," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. I–112–15, April 2003.

[6] Y. Agiomyrgiannakis and Y. Stylianou, "Coding with side information techniques for LSF reconstruction in Voice over IP," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 141–144, 2005.

[7] H. Ehara and K. Yoshida, "Decoder initializing technique for improving frame-erasure resilience of a CELP speech codec," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 549–553, April 2008.

[8] S. Praestholm, S.S. Jensen, S.V. Andersen, and M.N. Murthi, "On packet loss concealment artifacts and their implications for packet labeling in voice over IP," *IEEE International Conference on Multimedia and Expo*, vol. 3, pp. 1667–1670, June 2004.

[9] M. Jelinek, R. Salami, S. Ahmadi, B. Besette, P. Gournay, and C. Laflamme, "On the architecture of the CDMA2000 variable-rate multimode wideband (VMR-WB) speech coding standard," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. I–281–4, May 2004.

[10] T. Vaillancourt, M. Jelinek, R. Salami, and R. Lefebvre, "Efficient frame erasure concealment in predictive speech codecs using glottal pulse resynchronisation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. IV–1113–IV–1116, April 2007.

[11] V. Eksler and M. Jelinek, "Glottal-shape codebook to improve robustness of celp codecs," *IEEE Trans.on Speech and Audio Processing*, vol. 18, pp. 1208–1217, 2010.

[12] A.M. Gomez, J. L. Carmona, A.M. Peinado, and V. Sanchez, "A multipulse-based Forward Error Correction technique for robust CELP-coded speech transmission over erasure channels," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 1258–1268, 2010.

[13] Z.M. Lu, B. Yan, and S.H. Sun, "Watermarking combined with celp speech coding for authentication," *IEICET Trans. on Inf. and Systems*, vol. E88-D, no. 2, pp. 330–344, 2005.

[14] P. Vary and B. Geiser, "Steganographic wideband telephony using narrowband speech codecs," *Proc. Conference Record of Asilomar Conference on Signals, Systems, and Computers*, 2007.

[15] B. Geiser and P. Vary, "High rate data hiding in acelp speech codecs," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.

[16] F. Mertz and P. Vary, "Packet loss concealment with side information for voice over ip in cellular networks," *ITG- Fachtagung Sprachkommunikation*, 2006.

[17] A.M. Gomez, J. L. Carmona, J.A.Gonzalez, and V. Sanchez, "One-pulse FEC coding for robust CELP-coded speech transmission over erasure channels," *IEEE Transactions on Multimedia*, vol. 13, pp. 894–904, 2011.

[18] *Adaptive Multi-Rate (AMR);Speech processing functions;General description*, ETSI EN 301 703, 1998.

[19] *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assesment of narrowband telephone networks and speech codecs*, ITU-T P.862 Recommendation, 2001.

[20] J. Lindblom, J. Samuelson, and P. Hedelin, "Model based spectrum prediction," *in IEEE Workshop on Speech Coding*, 2000.

[21] R. Martin, C. Hoelper, and I. Wittke, "Estimation of missing LSF parameters using gaussian mixture models," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001.

[22] M. Serizawa and H. Ito, "A packet loss recovery method using packet arrived behind the playout time for CELP decoding," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. I–169–I–172, 2002.

[23] M. Chibani, R. Lefebvre, and P. Gournay, "Fast recovery for a CELP-like speech codec after a frame erasure," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2485–2495, Nov. 2007.

[24] V. Eksler and M. Jelinek, "Transition mode coding for source controlled CELP codecs," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4001–4004, April 2008.