PARTIAL SPECTRAL LOSS CONCEALMENT IN TRANSFORM CODERS

Anisse Taleb, Patrik Sandgren, Ingemar Johansson, Daniel Enström, Stefan Bruhn,
Multimedia Technologies, Ericsson Research, Sweden

ABSTRACT

In this paper, a novel error concealment technique for partial spectral loss in transform coders is presented. Based on amplitude and phase inter- and intra- frame correlations, an algorithm for missing spectral coefficients restoration is derived. The algorithm restores the missing spectral coefficients by predicting the amplitude using energy matching and the phase using group delay conservation principles. Results from listening tests illustrate the performance of the proposed algorithm.

1. INTRODUCTION

The transmission of compressed audio under adverse channel conditions is one of the most challenging problems in audio media distribution in wireless systems or the Internet. When all sender recovery schemes fail to correct the losses, error concealment at the receiver is often sought as the last resort in order to reduce the degradation of streamed audio quality [6].

Audio signals and especially speech signals often show large amounts of self-similarity. Error concealment techniques are based on producing a replacement for the lost packet, whose objective is the reduction of the perceived quality degradation. These techniques perform well for small packets and low loss rate (under 10%), for large packets the assumption of short-term self-similarity breaks down making these techniques of little use.
A large spectrum of error concealment techniques has been developed, for surveys see [6][7]. These include muting, repeating, noise injection, pattern matching, pitch based replication, spectral domain concealment [1], statistical restoration [2], etc.
In this paper, a transform based audio codec is considered. Without loss of generality, the transform domain will be referred to as spectral domain. The spectral coefficients corresponding to an audio signal frame are quantized and packetized for transmission. The problem considered concerns instances where only part of the spectral data relative to one audio frame is missing at the receiver side. In packet based transport this means that the encoded spectral coefficients corresponding to an audio frame are transported into different packets, which may be seen as an interleaving operation. Figure 1 illustrates an example of such spectral coefficients interleaving. This type of spectral interleaving appear for instance in the AMR-WB+ audio codec [8] where a super-frame of audio data is encoded and transported by 4 packets.

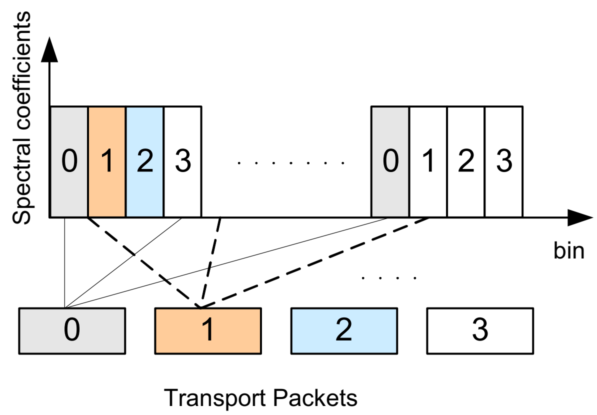

Figure 1: Spectral coefficients interleaving on 4 packets

Most existing concealment approaches, use only past and/or future information about the audio signal in order to conceal the packet loss. However, in the case of partial loss, some of the received coefficients relating to the same audio frame may be used in order to efficiently recover the missing coefficients. This type of concealment will use both inter and intra frame information.
The paper is organized as follows; in Section 2 a general description of the codec under consideration is given. Section 3 presents the principles underlying amplitude and phase concealment. The proposed algorithm is presented in Section 4 and in Section 5 listing test results illustrate the performance of the proposed algorithm.

2. DESCRIPTION OF THE SYSTEM UNDER CONSIDERATION

For the ease of development and without loss of generality, we consider a very basic frequency domain audio codec. The input signal is decomposed into overlapping frames where each frame is multiplied by a windowing function. The signal frames can be written as

$$x(m,n) = x(mK+n), n = 0, \Lambda\ , N-1$$

Where N is the window length, and $N-K$ is the length of the overlap, and the windowed signal frame is equal to:

$$x_h(m,n) = h(n)\cdot x(mK+n), n = 0,\Lambda\ , N-1$$

The resulting windowed signal frames are then transformed to the frequency domain by an FFT operation, after which the obtained spectral coefficients are quantized, packetized and transmitted.

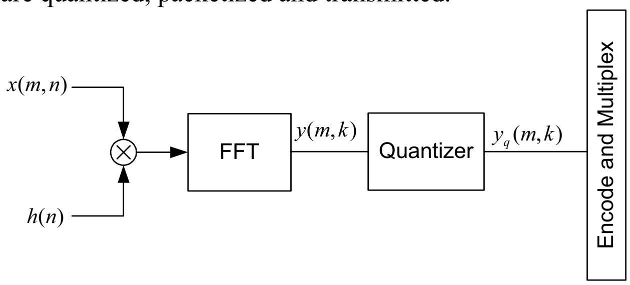

Figure 2: Basic frequency domain encoder

In the decoder side, the received spectral coefficients are first fed to an error concealment unit (ECU) the resulting coefficients are then transformed to the time domain by an inverse FFT operation. Finally, the obtained time-domain frames are overlap-added to produce the decoded audio signal.

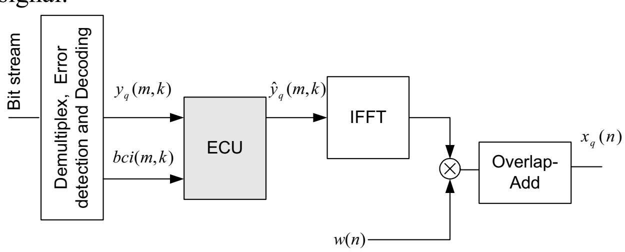

Figure 3: Basic frequency domain decoder

It is assumed the ECU receives an array of Boolean flags indicating which coefficients have been lost, which is a realistic assumption for packet based transport. Based on this information, the role of the ECU is to replace/repare the missing coefficients in order to decode the audio signal. When the ECU is not activated, the missing spectral coefficients are simply replaced by zeros, which can be seen as a muting technique.

3. AMPLITUDE AND PHASE CONCEALMENT

3.1. Experimental considerations

When no error concealment is used, replacing the missing spectral coefficients with zeros can be seen as a strategy for concealment. Muting the spectral coefficients, although very simple, does in fact lead to severe artifacts introduced by the large discontinuities it introduces. It is found experimentally that most annoying artifacts occur in audio frames that are stationary, e.g. long tones, voiced sections, etc. Packet repetition does not improve on muting; in fact repeating the previous spectral coefficients leads to severe artifacts due to phase mismatch on frame boundaries. This manifests itself as clicks, which occur in the overlapped sections of the audio signal frames.

Experimentally, it has been noticed that a simple scheme consisting of repeating the previous spectral coefficient amplitude leads to very good results provided that the phase of the missing spectral coefficient is correct, i.e. assuming that only the amplitude has been lost. The inverse experiment, where the phase is repeated while the amplitude is correct does lead to severe artifacts. This shows that phase concealment is much more demanding than amplitude concealment.

3.1. Group delay conservation property

Consider a single stationary complex tone

$$x(n) = A\cdot e^{jn\omega_0 + j\varphi_0}$$

The signal frames prior to frequency transform writes as

$$x_h(m,n) = A\cdot h(n)\cdot e^{j(mK\omega_0+\varphi_0)+j\omega_0 n}, n = 0,\Lambda\ , N-1$$

The Fourier transform of the signal frame writes as

$$X(m,\omega) = A\cdot H(\omega-\omega_0)\cdot e^{j(mK\omega_0+\varphi_0)}$$

The phase of the signal therefore writes as

$$\varphi_x(m,\omega) = \varphi_h(\omega-\omega_0) + mK\omega_0 + \varphi_0$$

It is well known that the temporal difference of successive phases is given by:

$$\Delta\varphi_x(m,\omega) = K\omega_0$$

This shows that the phase evolves linearly on each frequency bin. This property is exploited for instance in time-scaling/pitch shifting by the phase vocoder, where it is used in order to compute the phase of the scaled signal. It has also been exploited for transform coding of audio signals [6], where it is used in combination with a predictive phase quantizer. In [7], based on the same principle, a phase prediction is used for error concealment where past values belonging to the same frequency bin are used in order to predict the missing/erroneous spectral coefficient. This ensures that the phase will evolve coherently in time; however there is no guarantee that phase coherence is preserved in the frequency domain, i.e. with neighboring bins.

If one takes the derivative of the phase with respect to the frequency it leads to the group delay

$$\tau_x(m,\omega) = \tau_h(\omega-\omega_0)$$

This result shows that for a single stationary tone, the group delay is “frame” constant (independent of m). In particular, we have

$$\tau_x(m,\omega) = \tau_x(m-1,\omega)$$

It is easily shown that this result holds approximately true for multiple tones that are far enough from each other and is approximately locally true.

The group delay conservation property does relate the phase on a bin with the past phase of the bin and also the neighboring bins. Using this property in order to restore the phase ensures phase coherence for both time and frequency evolution.

4. SPECTRAL COEFFICIENT CONCEALMENT ALGORITHM

In the following we consider the following set of indices of the lost spectral coefficients,

$$S = \{k, bci(m,k) = TRUE\}$$

Each spectral coefficient is written in polar coordinates as

$$y(m,k) = Y(m,k)e^{j\varphi(m,k)}$$

where $Y(m,k)$, resp. $\varphi(m,k)$, denotes the spectral amplitude, resp. spectral phase, for frequency bin k at time frame m.

4.1. Amplitude concealment

As discussed earlier, amplitude concealment can be simply repeating the previous bin amplitude, i.e.

$$\hat{Y}_q(m,k) = \hat{Y}_q(m-1,k)$$

This can have the drawback that if the audio signal has a decreasing magnitude; over-estimation of the amplitude can occur. A more elaborate scheme uses intra-frame redundancies, which allows a better prediction of the spectral magnitude. The predicted spectral magnitude can be written as

$$\hat{Y}_q(m,k) = G(m) \cdot Y_q(m-1,k)$$

where $G(m)$ is an adaptive gain obtained by matching the energy of the non-lost spectral coefficients of the current frame with the corresponding spectral coefficients of the previous frame. An example of energy matching can be to compute the adaptive gain as

$$G(m) = \sqrt{\frac{\sum_{k \notin S} Y_q(m,k)^2}{\sum_{k \notin S} Y_q(m-1,k)^2}}$$

Other types of energy matching measures may be possible, for instance, the gain $G(m)$ can be estimated on several spectral bands by grouping the spectral coefficients into sub-bands and estimating the gain in each sub-band. The sub-band grouping can be on a uniform scale or a psychoacoustically motivated nonlinear frequency scale. The adaptive gain in sub-band l can therefore estimated by

$$G(m,l) = \sqrt{\frac{\sum_{\substack{k \in subband(l) \\ k \notin S}} Y_q(m,k)^2}{\sum_{\substack{k \in subband(l) \\ k \notin S}} Y_q(m-1,k)^2}}$$

And the predicted amplitude of the spectral coefficients in frequency sub-band l is given by

$$\hat{Y}_q(m,k) = G(m,l) \cdot Y_q(m-1,k), \quad k \in subband(l)$$

The estimated gain on each spectral band greatly benefit from smoothing both in the time domain (smoothing in m) as well as in the frequency domain (smoothing in l) by the use of, for example, low pass filtering in the time and the frequency domain or polynomial fit in the frequency domain and low pass filtering in the time domain.

4.2. Phase concealment

Phase concealment is based on the use of group delay conservation properties. A straightforward way consists of requiring that the concealed phase obey the group delay preservation. Phase prediction will then consist of first estimating the group delay from the previous frame, and then matching the current frame group delay to the previously estimated.

A simple way to estimate the group delay is to approximate the derivative of the phase in the previous frame by using first order finite differences, i.e.

$$\Delta\varphi(m-1,k) = \varphi(m-1,k) - \varphi(m-1,k-1)$$

The idea then is to restore the same group delay on all the missing spectral components. This can be achieved by computing the predicted phases such that they minimize an error function, for example:

$$\sum_{k \in S} W(k) \cdot (\varphi(m,k) - \varphi(m,k-1) - \Delta\varphi(m-1,k))^2$$

where the unknown parameters are $\varphi(m,k)$ such that k represents the bin of a missing spectral coefficient, while $W(k)$ are positive weighting coefficients.

It is advantageous that the weighting coefficients are set proportional to the magnitude of the spectrum of the previous frame, or the predicted magnitude of the current frame, or a smoothed spectral envelope. This has the effect of emphasizing the importance of the spectral peaks and to filter out the bad estimates of the phase derivative introduced by noise in the spectral valleys.

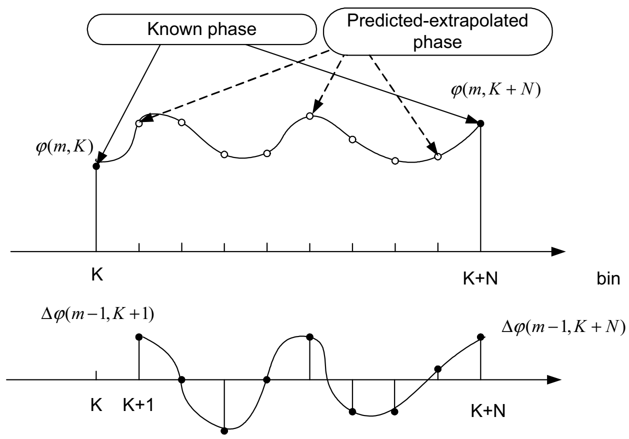

Figure 4: Illustration of the proposed concealment algorithm

An example of a solution in the case of $W(k)=1$ is given. As shown in Figure 4, the lost spectral coefficients are between bin K and bin $K+N$. The minimization of the error criterion leads to the following recursive solution for the extrapolated-predicted phase,

$$\hat{\varphi}(m,k)=\hat{\varphi}(m,k-1)+\Delta\varphi(m-1,k)+\Delta\varphi_c,$$
$$k=K+1,\ldots,K+N-1$$

where

$$\Delta\varphi_c=(1/N)\cdot(\varphi(m,K+N)-\varphi(m-1,K+N)$$
$$-\varphi(m,K)+\varphi(m-1,K))$$

and $\hat{\varphi}(m,K)=\varphi(m,K)$ is used to start the recursion.

The algorithm has an efficient iterative implementation that avoids complex trigonometric functions. This allows an application even under strict requirements on decoder complexity.

5. TEST RESULTS

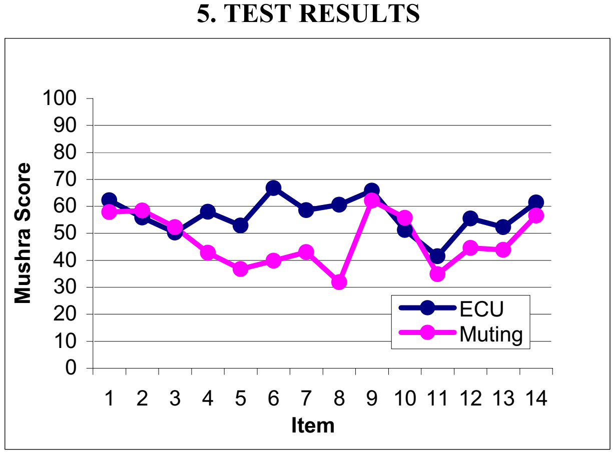

Figure 5: Mushra test results

Extensive listening test have been conducted during the the development of this algorithm. The results shown in the following figure show the improvement of the algorithm over a simple muting strategy in the case of 5% frame erasures rate. The large improvements were obtained on stationary music (items 4,5,6,7,8). These results show also how the algorithm succeeds in providing a consistent codec quality over a large variety of items.

The algorithm presented here is also part of the AMR-WB+ audio codec. Official test results performed by 3GPP during audio codec selection [3] show that the AMR-WB+ codec was the only codec to maintain a good quality (>50 Mushra points) even in impaired channel conditions.

5. CONCLUSION

In the case of partial spectral loss, error concealment should re-use as much information as possible from the currently received non-lost spectral coefficients in order to conceal the lost coefficients. In this paper a novel technique for error concealment has been developed that reuses past as well as presently received information in order to rebuild the missing part of the spectrum. The presented algorithm has been successfully implemented as part of the AMR-WB+ decoder.

Extensions to this algorithm for use with other types of transforms, e.g. DCT, are possible and are currently under investigation.

4. REFERENCES

[1] J. Herre, E. Eberlein, "Error Concealment in the spectral domain", 93rd AES Convention, 1992 Oct, 1-4, preprint 3364.

[2] R. Veldhuis "Restauration of lost samples in digital signals", Prentice Hall, 1990.

[3] S. R. Quackenbush, ARL. "Global Analysis Laboratory Report on 3GPP Low-Rate Audio Exercises", SA-040173.

[4] Y. Mahieux, "Error handling process for digital audio signal transmission system using Fourier transform to enable prediction of data, and concealment of signal errors" , France Telecom, EP-574288.

[5] Y. Mahieux, J. P. Petit, A. Charbonnier, "Transform coding of audio signals using correlation between successive transform blocks" ICASSP, May 1998, NewYork, pp. 2021-2024.

[6] C. Perkins, O. Hodson, V. Hardman, "A survey of packet loss recovery techniques for streaming audio". IEEE Network, Volume: 12, Issue: 5, September/October 1998. pp. 40-48.

[7] B. W. Wah, X. Su, and D. Lin. "A survey of error-concealment schemes for real-time audio and video transmissions over the internet". In Proc. Int'l Symposium on Multimedia Software Engineering, pages 17-24, Taipei, Taiwan, December 2000. IEEE.

[8] 3GPP TS 26.190, "3rd generation partnership project, technical specification group services and systems aspects, speech codec speech processing functions, AMR wideband speech codec; transcoding functions"