AN ATM SPEECH CODEC
WITH IMPROVED RECONSTRUCTION OF LOST CELLS

Kai Clüver

Institut für Fernmeldetechnik, Technische Universität Berlin
Einsteinufer 25, D-10587 Berlin, Germany
telephone: +49 30 314-24581; fax: +49 30 314-25799
e-mail: cluever@ftsu00.ee.tu-berlin.de

ABSTRACT

A speech codec for ATM networks is presented which includes ATM adaptation layer functions, a voice activity detection, and a new method for the reconstruction of lost cells. As the cell assembly already requires a relatively high buffering delay, only algorithms are applied which introduce small additional delays. The reconstruction of lost cells is based on an analysis of the LPC and pitch parameters of the speech signal. The new waveform substitution method considerably reduces the speech quality impairment caused by cell loss.

1. INTRODUCTION

The asynchronous transfer mode (ATM) to be introduced for future broadband integrated services digital networks poses problems to relatively low-rate, time-constrained services like duplex telephone speech. One problem is a higher transmission delay, compared with synchronous networks, which is caused by the buffering delay for the ATM cell assembly in the encoder and, to a lesser extent, by the buffering of the cells in the network nodes. An additional delay in the ATM decoder is necessary for the compensation of the network delay fluctuation.

Loss of ATM cells forms another problem of the asynchronous transfer mode. Cell loss is caused by the fluctuation of the network delay (if an ATM cell arrives late at the decoder, it has to be discarded) and by buffer overflows due to network overload. While the intelligibility of the speech signal is retained even at relatively high cell loss ratios [1], the perceived quality may be severely degraded. To prevent this, the decoder has to generate a substitute for the lost cell which reduces the perceptibility of the loss.

As the interpolation of adjacent valid cells for the reconstruction of a lost cell would require an additional delaying of the speech signal, only extrapolation of the speech signal preceding the missing cell is considered here. Simple methods, like repetition of the previous cell, yield only small improvements compared with silence substitution. Using more complex substitution techniques, the quality impairment caused by lost cells can be reduced. Usually, repetition of the preceding speech waveform at the pitch period is regarded as most efficient [2].

In this paper, the structure of a speech codec for ATM is described and an improved method for the reconstruction of missing cells is presented. Some results of a real-time implementation of the ATM codec are reported. Finally, the performance of the reconstruction method is discussed.

2. A SPEECH CODEC FOR ATM

The schematic diagram of a duplex speech transmission link through an ATM network is shown in fig. 1. A terminal adapter performs the ATM layer functions, including generation and extraction of the cell header, and links the speech codec to the ATM network. The ATM codec performs the speech-specific functions of the ATM adaptation layer (AAL), i. e., only the information fields of the ATM cells are processed by the codec. The information field format is of type 1 [3], resulting in 47 octets per cell available for the transmission of the coded speech. The speech codec employs G.711 logarithmic PCM with a bit rate of 64 kbit/s. Consequently, the buffering delay for cell assembly amounts to 5.9 ms.

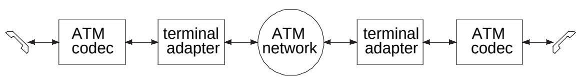

Fig. 1 Block diagram of the ATM speech transmission link

In the encoder, the input speech is buffered and every 47 samples the buffer contents is coded to 47 octets. The 4-bit sequence number is incremented and four protection bits are added to form the first octet of the information field. The entire information field of 48 octets is then passed to the terminal adapter for transmission. Furthermore, the speech encoder includes a voice activity detector which suppresses the transmission of ATM cells during silence periods and thus reduces the average network load. The decision on either voice activity or silence is done for each cell; it is based on an analysis of the speech signal already buffered for cell assembly. Thus, no additional signal delay is introduced. The buffering interval of 5.9 ms, however, is rather short for a reliable decision. Therefore, the decision has to be biased towards voice activity in order to prevent truncation of active speech. Consequently, the efficiency of the voice activity detector is somewhat reduced. Additionally, the beginning of each silence period is delayed by approximately 90 ms: a hangover of 15 consecutive ATM cells originally detected as silence are transmitted before the link is interrupted. The last cell of these is marked by a specific value of the sequence number to be utilized in the decoder.

The decoder reads the information fields of received cells from the terminal adapter and checks the sequence number. For valid cells, the octets containing the PCM data are decoded and 47 speech samples are played out. Silence periods are identified by the sequence number of the last transmitted cell. When such a marked cell is received, the coefficients for a linear predictive synthesis filter and the energy of the LPC residual are determined from an analysis of the decoded signal. During the silence period, a comfort noise signal is generated by exciting the synthesis filter with an appropriately scaled white noise signal. Assuming the decoded signal immediately preceding to contain no active speech, the comfort noise models the background noise in the transmitted signal.

Lost cells are reconstructed using an improved extrapolation method (see section 3). In case of late arrival, a cell is also treated as lost and the signal is extrapolated. At the succeeding cell interval, both the late cell that has meanwhile arrived and the now valid cell have been received by the terminal adapter. After reading both cells, the decoder identifies the late cell by its sequence number and subsequently discards it. Finally, the valid cell is decoded. For misinserted cells, a similar consideration applies. Therefore, the decoder reads two cells at a time, if present, from the terminal adapter in order to be able to identify late or misinserted cells.

3. RECONSTRUCTION OF LOST CELLS

The ATM decoder employs a reconstruction method for missing cells which forms an extension of the technique described in [2]. The new method further reduces the perceptibility of cell losses. It is based on the separation of the spectral envelope and the excitation signal by means of linear predictive (LPC) analysis. The structure of the reconstruction method is shown in fig. 2. For each valid cell, a set of predictor coefficients is calculated from the PCM decoded speech samples. These LPC coefficients are used to adapt two filters, a prediction error filter and a synthesis filter. The speech signal is passed through the prediction error filter and the resulting LPC residual is then used as excitation signal of the LPC synthesis filter. As the filters have mutually inverse transfer functions, the synthesis filter output is equal to the input speech.

In case of a cell loss, no LPC analysis is performed and the filter coefficients of the preceding cell are retained. The pitch period is determined by means of an auto-correlation analysis of a low-pass filtered version of the previous excitation signal. The reconstructed speech signal is obtained by repeating the LPC residual at the pitch period and then passing this substitute excitation signal through the synthesis filter. As the synthesis filter determines the spectral envelope of the output signal, variations of the speech spectrum during the reconstruction are thus prevented. Discontinuities at the boundary of the substitute excitation and the LPC residual of the succeeding valid cell are smoothed by merging both signals at the beginning of the next valid ATM cell. Furthermore, the reconstructed speech signal is fed back to the prediction error filter in order to retain the continuity of the filter memory.

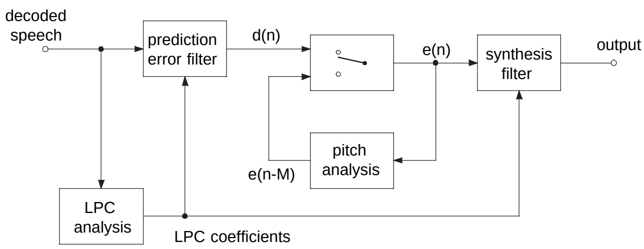

Fig. 2 Structure of the lost cell reconstruction algorithm

The reconstruction method works well for single or double cell losses. In case of bursty cell losses, however, the output signal contains tonal artifacts due to the periodical reconstruction in unvoiced or partly voiced speech. Consequently, a maximum of two consecutive missing cells are reconstructed as described above. For longer bursts, the substitute excitation is subsequently faded out and the decoder reverts to playing out comfort noise. For a smooth transition, both signals are merged at the beginning of the third missing cell.

4. EXPERIMENTAL RESULTS

A real-time version of the ATM speech codec was implemented on a prototype hardware containing two digital signal processors DSP32C. The delays for the various stages of signal processing are given in table 1. The codec showed a back-to-back delay of approximately 10 ms. This includes 1 ms to compensate for both the short-term jitter of the network delay and the longer-term fluctuation of the average delay due to changes in the background load of the ATM network.

The proposed reconstruction method was investigated for various channel conditions. Listening tests showed that most of the cell losses are rendered imperceptible at loss ratios of up to 10 %. Up to two consecutive missing cells can be reconstructed without adverse effects on the perceived speech quality. Fig. 3 shows a waveform example; in fig. 3(b), a missing cell was replaced with zero-valued samples. Fig. 3(c) depicts the waveform of the LPC residual signal, the missing cell now replaced by repetition of the preceding residual at the pitch period. The resulting synthesis filter output is shown in fig. 3(d).

buffer delay for cell assembly	5.9 ms
encoder computation time	0.5 ms
transfer to/from terminal adapter	0.4 ms
compensation of network delay fluctuation	1.0 ms
decoder computation time	2.5 ms
total	10.3 ms

Table 1 Signal delay of the real-time codec

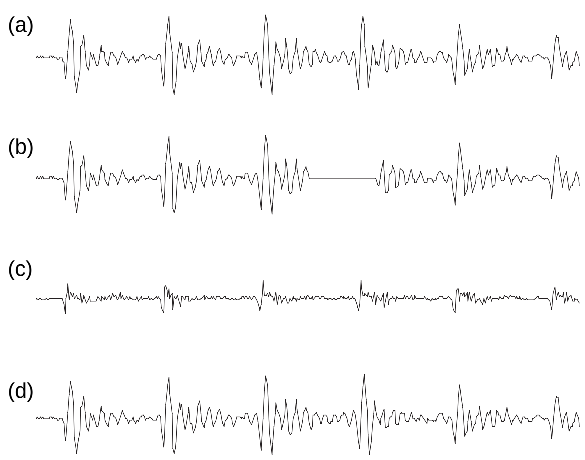

Fig. 3 Reconstruction of lost cells: (a) original speech waveform, (b) output waveform with silence substitution, (c) LPC residual with pitch waveform replication, (d) synthesis filter output

The performance of the reconstruction method was assessed in a formal listening test. Table 2 shows the mean opinion scores (MOS) for silence substitution and the proposed method. At a cell loss ratio of 1 %, the quality of the original speech signal is preserved. Above 5 %, the MOS ratings improve by about 1.4, compared with silence substitution. Additionally, informal listening tests showed that the proposed method outperforms the reconstruction of the speech waveform according to [2] (without LPC analysis).

cell loss ratio	silence	LPC/pitch
0 %	3.97	
1 %	3.58	4.07
5 %	2.42	3.81
10 %	1.78	3.33
20 %	1.17	2.67

Table 2 MOS results for silence substitution and the proposed reconstruction method

5. CONCLUSION

A speech codec for the asynchronous tranfer mode was presented which includes the ATM adaptation layer (AAL) functions, such as cell assembly and disassembly, detection of lost cells, and identification of late or misinserted cells. A voice activity detection reduces the average network load; during silence periods, a comfort noise signal is generated which models the background noise. Missing ATM cells are reconstructed by means of an improved method which is based on an analysis of the LPC and pitch parameters of the speech signal. The reconstruction method yields high quality gains under cell loss compared with simple silence substitution. Furthermore, it was found that *LPC residual* waveform substitution forms a clear improvement compared with *speech* waveform substitution. At a cell loss ratio as high as 10 %, a MOS of 3.3 is achieved, while the additional signal delay of the codec is kept at a minimum.

ACKNOWLEDGEMENT

This work was supported by Siemens AG, Munich, Germany.

REFERENCES

[1] A. Hills, K. Scott: Perceived Degradation Effects in Packet Speech Systems, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-35, no. 5, pp. 699-701, May 1987

[2] O. J. Wasem, D. J. Goodman, C. A. Dvorak, H. G. Page: The Effect of Waveform Substitution on the Quality of PCM Packet Communications, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 36, no. 3, pp. 342-348, March 1988

[3] CCITT: ATM Adaptation Layer (AAL) Specification, Recommendation I.363, Geneva 1991