ON THE COMBINATION OF REDUNDANT AND ZERO-REDUNDANT CHANNEL ERROR DETECTION IN CELP SPEECH-CODING

Norbert Görtz
Institute for Network and System Theory
University of Kiel, Germany
Tel.: +49 431 77572 406, Fax: +49 431 77572 403
e-mail: ng@techfak.uni-kiel.de

ABSTRACT

In this paper a new algorithm is described for selective detection and handling of speech codec parameters which were corrupted by bit-errors on the transmission channel. A combination of classical forward error detection schemes using additional (redundant) bits and parameter-correlation based techniques without redundant bits (zero-redundant error detections) is used for this purpose. The algorithm is optimized by informal listening tests rather than by maximization of mathematically tractable measures (e.g. SNR) which usually do not reflect subjective speech quality well. No additional delay and almost no additional memory and complexity is required for the new algorithm. The speech quality resulting at the decoder output is strongly improved compared to standard bad-frame handling techniques if coded speech is transmitted over disturbed channels, e.g. the GSM-full-rate channel which is used for performance evaluation.

1 INTRODUCTION

In the implementations of presently used speech codecs, e.g. the full-rate GSM-speech-codec, uncorrected channel errors are detected by classical forward error-detections like Cyclic Redundancy Checks (CRC). Because of bit-rate limitations only a few bits are available for the CRCs, so error detection must be carried out commonly for a *set of parameters*. In case of an error detection the whole set is replaced by extrapolated values derived from uncorruptedly received parameters from previous frames by exploiting the parameter-correlation in time. Thereby parameters which were not corrupted are also replaced, so correctly received information is "thrown away". The output speech quality of the decoder would be better, especially on highly disturbed channels, if only those parameters were replaced which are corrupted.

The paper is organized as follows: In section 2 a codec developed for enhanced speech-transmission in the GSM-system [5] is briefly described. It is used as a reference codec for the following investigations. Thereafter, time-based and mutual dependencies of the codec parameters are stated in section 3, and they are exploited for parameter extrapolation and channel-error detection in sections 4 and 5. Then the zero-redundant error detections are combined with forward error detections in section 6. Finally the performance of the proposed algorithm is discussed in section 7.

2 SPEECH CODEC

The CELP encoder processes frames of 160 narrow-band speech samples (20 ms), which are divided into 4 subframes. Figure 1 shows the block diagram of the encoder. The total number of bits per frame is 214, i.e. the bit rate is 10.7kbps.

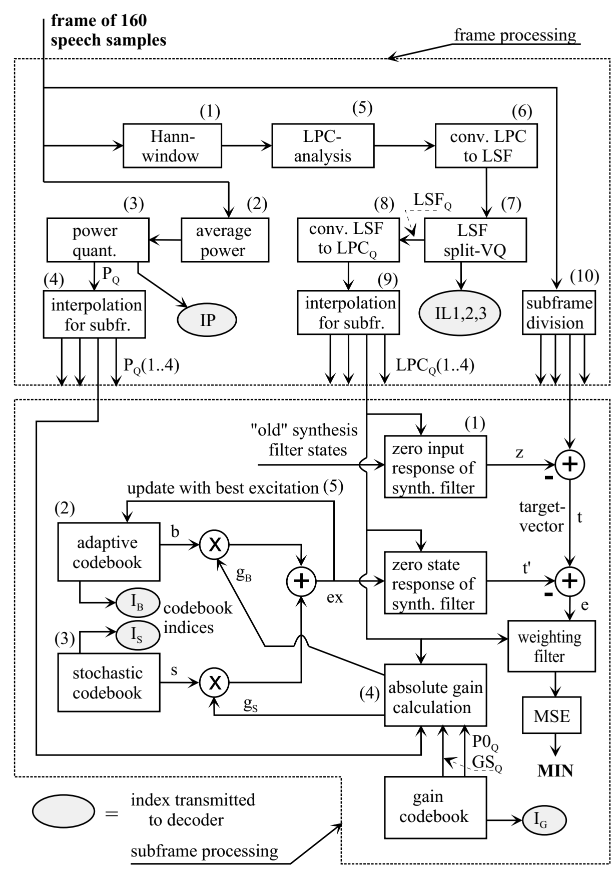

Figure 1: Speech Encoder

2.1 Frame Processing

First, the average power of the signal is calculated. It is logarithmically quantized by a 5-bit table, similar to [2]. Then a 10^{th}-order LPC analysis is carried out. The LPC coefficients are converted to Line Spectrum Frequencies (LSF) using the algorithm in [3]. They are split-vector-quantized according to the method in [4], resulting in 3 codebook indices for the 10 LSF, with 9+8+8=25 bits.

2.2 Subframe Processing

The search for the best components of the excitation vector ex, which is the input signal of the LPC-synthesis filter, is performed sequentially with the adaptive-excitation vector b first. While searching for the best vectors b and s out of the codebooks the optimal unquantized gains are used. When the best adaptive- and stochastic-excitation vectors have been found, the best gain codevector for those excitation vectors is searched "closed-loop" in the gain codebook.

The adaptive codebook is searched "closed-loop" over a "lag"-range of 20..141. The best lag (number of samples back in the past where the adaptive excitation starts) is coded by 8 bits (index I_B). Up to 5 non-integer values (fractional lags) between two integer samples are possible.

The stochastic excitation consists of ten +1/-1 pulses which are systematically placed in the excitation vector and coded by 30 bits (index I_S), similar to ACELP [6].

The excitation signals are scaled by two gains which are jointly quantized by an 8 bit codebook trained by the LBG-algorithm [7] similar to the method in [2]. The components of the codebook are $P0$, the power of the adaptive excitation divided by the sum of the powers of the excitation signals, and GS, which compensates for the error in the estimation of the excitation signal power by the sum of the powers of the components b and s, neglecting their correlation.

2.3 Decoder and Postfiltering

As usual in "analysis-by-synthesis" codecs, the operations to be performed in the decoder (except post-processing) are similar to those already carried out in the corresponding encoder stages. A postfilter is employed to increase the speech quality at the decoder-output in terms of human perception. It includes long-term and short-term filtering similar to [8].

3 PARAMETER-DEPENDENCIES

A large set of speech data (100000 frames) was coded, and the quantized parameters of the speech codec, i.e. the average power, the LSF-coefficients, the lags, and the gains, were used to calculate the probability distributions of the differences of adjacent parameters with a distance of D frames/subframes. Some results are plotted in figure 2. Figure 2 reveals that the log. average power logP, the LSF-coefficients (only the first LSF-coefficient is plotted), the lag and the normalized power P0 of the adaptive excitation have a considerable correlation in time, i.e. the differences of adjacent parameters tend to be smaller if their distance D is small. For $D > 4$ frames/16 subframes (80msec) the parameters are almost independent. The reason for this can be derived from a time-domain plot of a speech signal: Most of the phonemes, which are the "areas of similarity", don't last longer than 80ms.

Beyond the correlations in time, the differences of consecutive lags are correlated with P0: If P0 is close to 1, the speech signal is often voiced and the lag-differences are small with higher probability than given by figure 2. In addition to that, the lag values in voiced subframes correspond to the basic pitch period or integer multiples or parts of it. Therefore the differences of consecutive lags are even more limited if lag-values are considered that are corrected to the basic pitch period. Also the LSF-coefficients *within* a frame are correlated: The differences of LSFs with neighbouring indices only have one sign and their absolute values are limited too. Further details were published in [5].

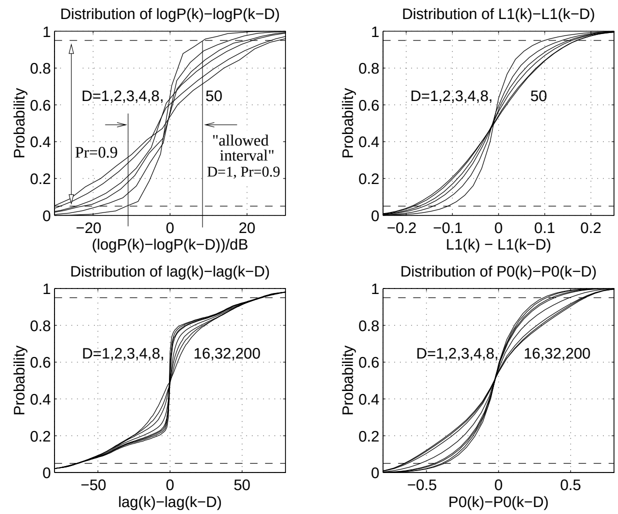

Figure 2: Probability distributions of differences of the quant. parameters with a distance of D frames/subframes

4 BAD-PARAMETER HANDLING (BPH)

If a parameter corruption has (somehow) been detected, bad-parameter handling is carried out in the speech decoder by replacing the corrupted parameter of the current frame by the last uncorrupted parameter from a previous frame. This basic idea is justified by the correlations in time measured in section 3. For situations with consecutive parameter-corruptions, the last uncorrupted parameters are modified before use: The power, for example, is decreased with the number of adjacent corrupted indices to muffle the extrapolated signal and thereby to avoid annoying distortions in the output signal. The LPC-poles are radially shifted towards the origin of the Z-plane, so peaks in the spectrum will more and more be flattened if several adjacent errors occur.

In the subframes, the extrapolation of corrupted parameters can be performed by exploiting uncorrupted parameters in *future* subframes within the frame, without adding additional delay to decoding. For the gains, linear interpolation with the last and the future uncorrupted values is carried out. For the lags the last uncorrupted value is used, if its distance D to the corrupted lag is smaller than the distance to a future uncorrupted value or if there is no future uncorrupted value within the frame. Otherwise the future uncorrupted lag-value is used for extrapolation.

5 ZERO-REDUNDANT ERROR DETECTION

In figure 2 dashed lines were printed into each of the subplots, corresponding to parameter intervals on the x-axes with the probability of 0.9. The intervals can be found by the intersections of the dashed lines with the probability distributions for each D as shown in the upper left plot in figure 2 for D=1. The resulting intervals are not exceeded by uncorrupted parameters with a probability of 0.9, i.e. *if* they are exceeded this is caused with a probability of 0.9 by a channel error. So, parameter corruptions by bit-errors can be detected by checking, whether the intervals are exceeded. Since the probability of the interval was 0.9, there will be "wrong" error detections with a probability of 0.1 resulting in unnecessary bad parameter handling and consequently in reduced clear channel speech quality. Since no additional (redundant) bits are used for error detection it is called "zero-redundant" although the redundancy of the parameters (correlation in time) is exploited.

The interval-probabilities of the zero-redundant error detections can be optimized by listening tests so speech quality is clearly better (no "clicks" and "plops") on channels with bit errors while the clear channel quality is only moderately reduced. For details the reader is referred to [5].

The zero-redundant error detections require only a few memory locations to store the interval-limits, e.g. 8 memory words for 4 adjacent realizations of a frame-parameter and 32 for a subframe-parameter to cover a time range of 80 ms. They do not cause significant computational load because the detections can be implemented by two simple comparisons of the interval-limits with the realization of a parameter difference once per frame/subframe.

6 COMBINATION OF FORWARD AND ZERO-REDUNDANT ERROR DETECTION

6.1 Forward Error Control

The error-patterns for the GSM full-rate channel are used to simulate a realistic "transmission environment". As usual in GSM, rate 1/2 convolutional coding is used for forward error correction. In contrast to the full-rate specifications, all of the bits of the speech codec are protected by the convolutional code because investigations of the bit-error sensitivity of the coded parameters revealed that there are no bits to be transmitted without protection on a channel with bit error rates up to 10%. Since the bit rate for source and channel-coding on the GSM full-rate channel is 22.8 kbps, there are 456 Bits per frame to be transmitted. The convolutional code has 456/2-4=224 Bits/Frame as input bits (4 zero-tail-bits are required at the end of a block to erase the memory of the convolutional encoder). The speech codec produces 214 Bits/Frame, so 224-214=10 Bits are left for additional forward error detection "inside" the convolutional code, where the remaining bursty bit-errors have error-rates up to 4.5%. Listening tests and error-detection statistics revealed that a minimum of 5 bits for a CRC must be used, to ensure safe error detection. After listening tests it was decided to check each of the frame and subframe parameters within a frame, by an own 5-bit CRC. Since parameter replacement (section 4) in case of bit errors is only worthwhile if the speech quality at the decoder-output is better than without it, the bits of the indices meeting this requirement were determined by informal listening tests. These bits (MSBs) are the ones which have to be checked by a CRC. In the left 4 columns of table 1 the forward error-detection scheme is summarized, and a block diagram for the frame parameters is displayed in figure 3. The digital substitute channel replaces the convolutional coding/decoding and all of the following steps of transmission. It is realized by error patterns.

6.2 Combined Error Detection

The forward and zero-redundant error detections are combined by logical AND. In figure 4 a block diagram of error detection and handling is shown for the first CRC and the mean power. For the other frame parameters the same CRC is used as forward error detection while each parameter has its own zero-redundant detection combined with the CRC in the same manner as for the mean power. For the subframe parameters the same structure is used, but the second CRC performs forward error detection. The outputs $P''_Q, LSF1''_Q, \dots$ of the error detection and handling are decoded by the standard process of the source codec, so the error-detector/handler in figure 4 is a preprocessor for the received indices/quantized parameters.

power	LSF1..3	LSF4..6	LSF 7..10	lag	gains
0.6	0.6	0.6	0.6	0.4	0.5

Table 2: Optimal interval probabilities for zero-redundant error-detections combined with forward error detections

By the combination of the zero redundant with forward error detections a loss of clear-channel speech quality can be avoided, since a CRC will never detect an error if there is no true error. Therefore the interval probabilities of the zero-redundant detections can be low; so the "allowed" intervals of parameter differences (refer to figure 2) are small, and more errors are discovered without affecting clear channel quality. The interval probabilities have to be optimized for the structure of the combined detection, i.e. how many and which parameters are checked by one CRC.

6.3 Optimization and Error Detection Statistics

For the optimization and evaluation of the performance of the new error detection scheme the worst error pattern available for the GSM full-rate channel (EP3 with Carrier to Interferer ratio C/I=4dB) was used. The results are first given as error-detection statistics in table 1. The resulting speech quality is discussed later. The CRCs detect about 96 % of the true errors in "important bits" (MSBs) in each parameter-index (refer to the right one of the "CRC" columns in table 1). Since the 4 indices of the frame parameters are checked by one CRC, there are many more error detections for each parameter than truly occur (refer to the left of the CRC columns in table 1). For the mean power roughly 4 times more error detections than true errors in the MSBs occur. Overall the detection-rates of the CRCs are in the range of 1.3 to 4.3 times the real errors for a single parameter.

The interval probabilities of the zero-redundant error detections for each parameter (combined with the CRCs) were optimized by informal listening tests. Ideal error detections based on MSB-comparisons of the corrupted and the uncorrupted indices were carried out for the parameters with error detections which were *not* in the optimization process. The optimal choices for the interval-probabilities (related to limits for parameter differences by the probability distributions in figure 2) are given by table 2. In table 1 the detection-statistics for the optimized zero-redundant error-detections are given in the two right columns. Between 91% and 94% of the true errors in the parameters are detected, but 7 up to 16.8 times more detections than true errors occur because the "allowed limits" of the differences of adjacent parameters (with a distance of D frames/subframes) are very tight (because of the low interval-probabilities in table 2). The combined detections (columns CRC $\wedge$ zero-red.) discover 90% ... 92% of the true errors and the total number of detections can be reduced (compared to the CRC-only detection) to $1.2 \dots 2.9$ times of the true errors. The number of detections for the mean power can be reduced from 429% to 292% of the true errors, so many bad-parameter handlings caused by the CRC alone will not be carried out since the combined detection did not find an error. Unfortunately there are true MSB-errors which are not detected by the combined technique, so bad-parameter handling is *not* carried out and the MSB-corrupted index is used for speech decoding. This is

parameter	index	MSBs	CRC-number	number of true errors in MSBs	detected errors in % related to true errors in MSBs: CRC ∧ zero-red. all	CRC ∧ zero-red. true	CRC all	CRC true	zero-redundant all	zero-redundant true
mean power	IP	3..5	1	167	292.22	90.42	429.34	97.01	1678.44	93.41
LSF1..3	IL1	4..9	1	371	160.92	89.76	193.26	96.23	953.64	92.18
LSF4..6	IL2	3..8	1	484	125.21	89.67	148.14	95.45	702.69	92.15
LSF7..10	IL3	3..8	1	553	116.27	91.32	129.66	95.84	669.26	93.67
lag 1..4	IB	2..8	2	2120	163.58	89.20	197.36	97.59	860.85	91.23
gains 1..4	IG	4..8	2	2182	160.91	92.48	191.75	97.80	749.95	94.36

Table 1: Detection-statistics for GSM error-pattern "EP3" with $C/I = 4$ dB with 6000 GSM-frames, using 2 5-bit CRCs.

not a real problem as relatively small parameter differences result from these undetected errors (because of the low interval probabilities of the zero-redundant part of the detection), so that the use of corrupted parameters is similar to the replacement by the "old" values by bad parameter handling.

7 PERFORMANCE

The "clicks" and "plops" known from standard GSM-implementations have completely been removed. The speech signal at the decoder output contains only slight noise and roughness even under adverse transmission conditions, e.g. C/I=4dB. The well known "muting" of the signal caused by less selective bad-frame replacing techniques occurs only at much higher bit-error-rates.

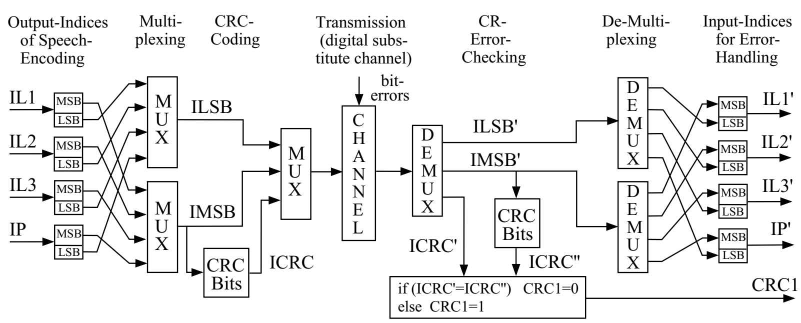

Figure 3: Forward error detection and transmission of the frame parameters

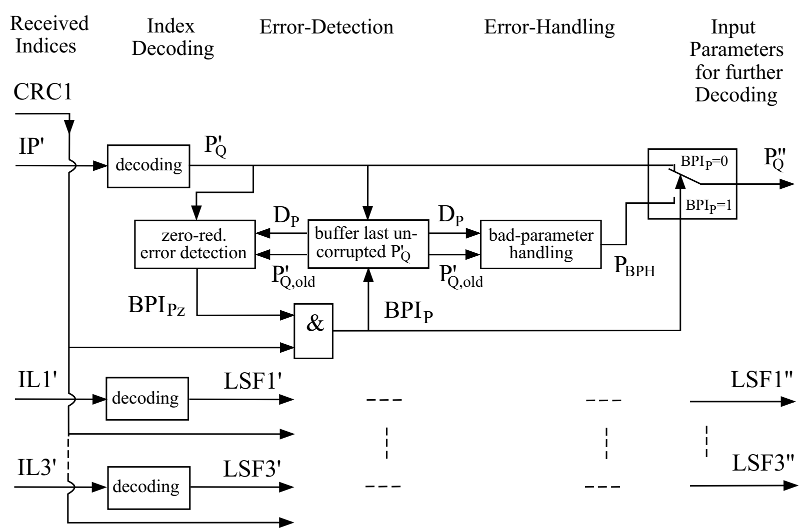

Figure 4: Combined error detection and bad parameter handling for the frame parameters.

8 CONCLUSIONS

An new algorithm for selective detection and handling of bit-errors in speech codec parameters has been stated. It does not require large amounts of memory since only small tables with the allowed-interval limits of the parameter differences have to be stored. Also additional complexity is not entailed since the zero-redundant error detections can be implemented by simple comparisons once per frame/subframe. Additional CRCs that improve selective error detection require only very small complexity and bit-rate. The speech quality under bad channel conditions has been strongly improved compared to standard frame-replacing techniques with less selective error detection.

9 REFERENCES

[1] N. Görtz, "Vorschlag für einen GSM-Vollraten-Sprach- und Kanalcodec mit verbesserter Sprachqualität", Proc. of "Aachener Kolloquium Signaltheorie 1997" (in German)

[2] Electronic Industries Association, Cellular Systems Dual-Mode Subscriber Equipment - Network Equipment Compatibility Specification, December 1989 (IS-54 US-standard)

[3] F.K. Soong, B.H. Juang, "Line Spectrum Pair (LSP) And Speech Data Compression", Proc. ICASSP 1984, pp. 1.10.1-1.10.4

[4] K.K. Paliwal and B.S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", IEEE Trans. on Speech and Audio Processing, Vol. 1, No. 1, pp. 3-14, January 1993

[5] N. Görtz, "Zero-Redundancy Error Protection for CELP Speech Codecs", Proc. EuroSpeech '97, Vol. 3, pp. 1283-1286, September 1997

[6] J-P. Adoul, P. Mabilleau, M. Delprat and S. Morisette "Fast CELP coding based on algebraic codes", Proc. ICASSP 1987, pp. 1957-1960

[7] Y. Linde, A. Buzo, R.M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. on Comm., Vol. COM-28, No. 1, pp. 84-95, January 1980

[8] ITU-T Rec. G.728, "Coding Of Speech At 16kbit/s Using Low-Delay Code Excited Linear Prediction", 1992