THE INFLUENCE OF TRANSMISSON ERRORS ON THE PERFORMANCE OF A CELP SCHEME

P. Siebert, H. Reininger, and D. Wolf

Institut für Angewandte Physik der Universität Frankfurt a.M.
D-6000 Frankfurt a. M., Robert-Mayer-Straße 2-4, FRG

RESUME

On présente l'influence des erreurs de transmission sur les paramètres d'un codeur CELP à 4.8 kbit/s. Les erreurs aléatoires aussi que les erreurs de burst sont considerées. La sensibilité des paramètres d'un prédicteur à court et à long terme, du gain et de l'excitation par rapport aux erreurs sur le canal de transmission est examinée. Les résultats montrent que surtout des erreurs d'estimation des paramètres du prédicteur à long terme et/ou le facteur d'amplification influencent la qualité de la parole.

SUMMARY

The influence of transmission errors on the parameters of a CELP codec for a transmission rate of 4.8 kbit/s was studied. Random as well as burst errors were considered. The sensitivity of the short-term predictor parameters, of the long-term predictor parameter, of the gain factor, and of the innovation sequence was examined. The results show that incorrect values of the long-term predictor parameters and/or the gain factor predominantly affect the speech quality.

1. Introduction

A promising method for encoding speech at low bit rates is the CELP-scheme [1,2]. CELP is based on the principle of analysis by synthesis as shown in Figure 1.

A speech signal $\{s(n)\}$ is processed in segments $s(n)$ of L sample values each. Each segment $s(n)$ is approximated by the output $\tilde{s}(n)$ of two time-varying digital filters $1/A(z)$ with

$$A(z) = 1 - \sum_{k=1}^{m_p} a_k \cdot z^{-k} \qquad (1)$$

and $1/P(z)$ with

$$P(z) = 1 - \sum_{k=-j}^{j} b_k \cdot z^{-(kp+k)}, \quad j = \frac{m_{pp}}{2} - 1 \quad . \qquad (2)$$

The filters are excited by the sequence $i(n)$ consisting of the components of an innovation vector $\underline{i}_j$ multiplied by a gain factor G_j. The vector $\underline{i}_j$ is selected from a codebook containing K vectors of dimension L. The difference $d(n)$ between $s(n)$ and $\tilde{s}(n)$ is weighted by the filter $W(z)$ and the mean squared error of the weighted sequence $w(n)$ is calculated. This procedure is repeated with all codebook vectors. The index k of that codebook vector yielding the smallest error is transmitted to the receiver together with the corresponding gain factor. The filter coefficients of $A(z)$ and $P(z)$ are encoded and transmitted to the receiver, too. In the receiver the speech segment $\tilde{s}(n)$ is reconstructed by exciting the two filters by the optimum codebook vector $\underline{i}_k$ scaled by the gain factor G_k.

The speech signal at the decoder may be disturbed in two ways. First quantization noise will be introduced by source coding. These effects have been studied elsewhere [3,4]. Second the transmitted data can be affected by channel errors. In this paper the influence of random and burst-like channel errors on the performance of the CELP scheme with a data rate of 4.8 kbit/s is examined.

2. The 4.8 kbit/s CELP-codec

For encoding the innovation a codebook with K=256 vectors of dimension L=40 was chosen resulting in a rate of 0.2 bit per sample or 1.6 kbit/s. The gain factor was quantized by an optimum scalar quantizer with 16 quantization levels. Thus, for encoding the innovation a total rate of 2.4 kbit/s is required. A long-term predictor with $m_{pp}=1$ was adapted every $L_{pp}=40$ samples [5]. The coefficient of the long-term predictor was quantized using an optimum scalar 3 bit-quantizer. The range of the delay kp was restricted to values between 40 and 71 resulting in a rate of 5 bits. For encoding the parameters of the short-term predictor the remaining 0.8 kbit/s of 4.8 kbit/s were used. They were quantized by a VQ scheme using a product codebook (PCVQ) composed of two codebooks of dimension m_1 and $m_2=m_p-m_1$. The short-term predictor was adapted every 160 samples. The order was set to m=8 and the dimensions of the two codebooks to $m_1=m_2=4$. The first codebook for encoding the first m_1 predictor parameters was tree structured with 2 levels and 32 branches per level. The first level of the tree search codebook provided a coarse quantization of the predictor parameters which is refined in the second level. Thus the computational cost was reduced drastically and the data were separated into more and less significant bits during the encoding process. The second codebook of the PCVQ scheme was a full search codebook with 64 vectors leading to a total rate of 16 bit per parameter vector. The bit allocation of this 4.8 kbit/s-codec is summarized in Table 1.

Table 1: Bit allocation of a 4.8 kbit/s CELP-codec

	Adaptation Period	Code Rate	Data Rate
	ms	bit	kbit/s
Excitation	5	8	1.6
Gain Factor	5	4	0.8
Short-Term Predictor	20	16	0.8
Long-Term Predictor	5	3	0.6
Pitch Period	5	5	1.0

With these parameters our CELP scheme achieved a SNR value of 12.2 dB. In informal subjective listening tests the speech quality was judged with scores between fair and good.

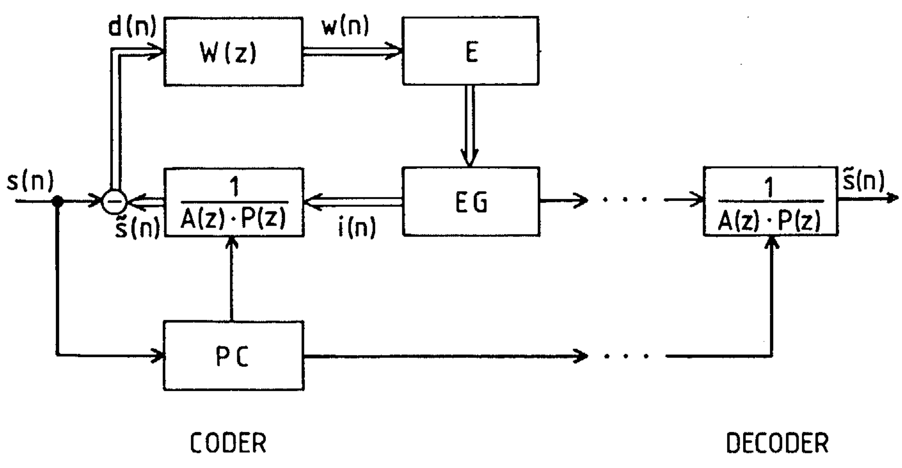

Figure 1 : Structure of the CELP codec

3. Simulation of the Transmission Channel

In order to study the influence of transmission errors on the CELP scheme noisy channels with statistically independent and with burst-like errors were simulated. The sequences of independent bit errors were generated by a random number generator producing uniformly distributed random numbers. If a random number of this generator was above a certain value LE an error bit was emitted. By appropriate choice of LE a desired error probability could be obtained.

Burst errors were produced by means of a two states channel model consisting of a state B with high error probability P_e^B and a state G with a very low error probability P_e^G. The transition between the two states was controlled by the crossing event of a stationary Rayleigh-process with respect to a certain level R_c. R_c was adjusted such that a given total bit error probability P_E was obtained. The different bit error sequences in the two states were generated by random number generators as described above [6].

4. Experimental results

The influence of transmission errors on the performance of the 4.8 kbit/s CELP scheme was analyzed in computer simulations using speech data of German male and female speakers. All utterances were low-pass filtered to 3.4 kHz and sampled at 8 kHz with a resolution of 16 bit per sample. As an objective performance measure the ratio between the power of the speech signal {s(n)} and the difference signal {d(n)}

$$SNR = 10 \log \frac{\sum_j \underline{s}_j \cdot \underline{s}_j^T}{\sum_j (\underline{s}_j - \tilde{\underline{s}}_j)\cdot(\underline{s}_j - \tilde{\underline{s}}_j)^T} \qquad (3)$$

was used. In addition the performance was evaluated by subjective comparative listening tests.

Introducing channel errors with an error probability of 1% during the transmission of the coded parameters from the sender to the receiver the performance of CELP deteriorates drastically. The SNR values given in Table 2 decreased and also the synthesized speech signal sounded severely distorted. This effect was more annoying in the case of the random errors as compared to burst errors due to the increased number of disturbed vector indices. However, in all cases a total failure of the CELP-codec was never observed indicating that error propagation in the decoder is limited.

In order to find out the influence of errors on the coder performance with a specific parameter on the coder performance only that parameter was left unprotected while the errors of all the other parameters were corrected. The measured SNR values are presented in Table 2. Disturbing the parameters of the short-term predictor led to a decrease in SNR of about 5 dB for the Rayleigh channel and 7 dB for the Random channel. In both cases the speech was found to be distorted by bubbling noise. Incorrect coefficients of the long-term predictor resulted in a drastic reduction in SNR and also in the speech quality. Some speech segments were completely unintelligible. Transmission errors in the delay kp of the long-term predictor reduced the SNR value only slightly. Finally, the innovation sequence was left unprotected. Incorrect gain factors decreased the SNR values drastically. The intensity of the signal at the receiver changed abruptly. The effect of transmission errors on the index of the innovation vectors were found to be of minor influence on the performance. The SNR decreased only slightly; the speech sounded somewhat rough. All these effects were more significant in the case of the random channel errors than in the case of the burst-like channel errors.

Table 2: SNR values after transmission over noisy channels with 1% error rate

Unprotected parameter	Random channel	Rayleigh channel
none	12.2	12.2
all	-4.3	0.1
Short-Term Predictor	5.1	7.3
Long-Term Predictor	-2.3	2.6
Pitch Period	5.2	7.5
Innovation		
Gain	2.9	5.3
Vector Index	9.8	10.8

The results show that especially the coefficient of the long-term predictor and the gain factor must be protected against channel errors. In order to reduce the amount of data which must be protected against channel errors it is important to determine the most sensitive bits. Therefore, the performance of the CELP-scheme was evaluated under the condition that one specific bit of a certain parameter codeword was permanently disturbed, while all of the other bits were transmitted correctly. This means that in the case of the short-term predictor parameter 1 bit of the 96 bits in each of the frames and in all other cases 4 bits of 96 bits were erroneous. The resulting SNR values are presented in Figure 2. Due to the high error rate in general the SNR values are reduced drastically. The amount of reduction indicates the influence of the specific errors on the performance. Obviously, erroneous long-term predictor coefficient and gain factor lead to severe degradation, whereas the influence of errors in all the other parameters on the performance is relatively small. Comparative listening tests confirmed these statements but also revealed that errors of the short-term predictor parameters were found to be more unpleasant than errors in the innovation vector index. In particular, disturbing less significant bits of the short-term predictor parameters affected the speech quality as much as errors in more significant bits.

5. Conclusion

The results show that in order to achieve at least a fair speech quality in the presence of channel errors the long-term predictor coefficient and the gain factors must be protected carefully. Especially, it is important to protect the two most significant bits of the long-term predictor coefficient and the 3 most significant bits of the gain factor. Erroneous bits in the other parameter codewords lead only to a slight degradation of the

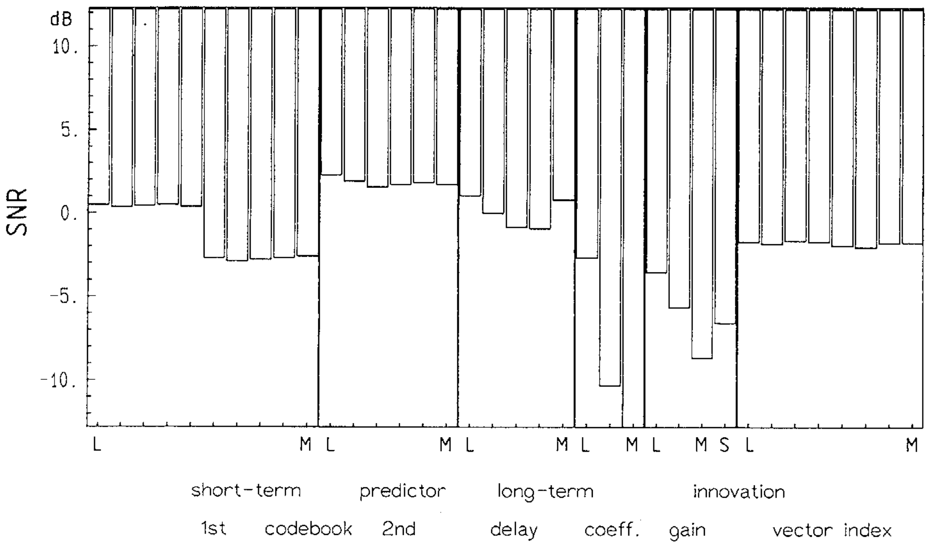

Figure 2 : SNR values due to single disturbed bits. Most (M) and least (L) significant bit, sign (S)

speech quality. These statements are in accordance with the observations of subjective listening tests. Thus, at least 1000 out of the 4800 transmitted bits per second must be protected by means of channel encoding. By additionally protecting the 800 bits of the short-term predictor parameters a speech quality was found which approaches closely that of undisturbed speech.

6. References

[1] Schroeder, M.S. and Atal, B.S., "Code-Excited Linear Prediction (CELP): High Quality at Very Low Bit Rates", Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, pp. 937-940 (1985)

[2] Buzo, A., Gray, R.M., and Markel, J.D., "Speech Coding Based Upon Vector Quantization", IEEE Trans. on Acoust., Speech, and Signal Processing ASSP-28, (1980) pp. 562-574

[3] Kroon, P. and Atal, B.S., "Quantization Procedures for the Excitation in CELP Coders", Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, pp. 1649-1652 (1987)

[4] Siebert, P., Reininger, H., and Wolf, D., "Quantization of Predictor Parameters and Gain Factors in a CELP Codec for a Transmission Rate of 4.8 kbit/s", Proceedings of EUSIPCO-88, pp. 1027-1030 (1988)

[5] Reininger, H., Siebert, P., and Wolf, D., "The Efficiency of Long-Term Predictors in CELP Schemes", Proc. of SPEECH'88, pp. 1033-1038 (1988)

[6] Reininger, H., "Prinzipien der digitalen Sprachcodierung und ihre Anwendung zur Sprachübertragung über Fadingkanäle bei mittleren Datenraten", Thesis Universität Frankfurt 1987