AN ALGORITHM FOR THE TRAINING OF CELP EXCITATION CODEBOOKS

Ulrich Balss, Herbert Reininger, Holger Schalk, Dietrich Wolf
Institut für Angewandte Physik der J.W. Goethe-Universität Frankfurt a.M.
Robert-Mayer-Straße 2-4, D-60054 Frankfurt am Main, FRG
Tel: +49 69 798 28163; Fax: +49 69 798 28510
e-mail: balss@apx00.physik.uni-frankfurt.de

ABSTRACT

CELP schemes with trained excitation codebook are able to reproduce more complex waveforms than stochastic CELP schemes. Here we present a new algorithm for the design of trained CELP excitation codebooks which are well adapted to the residual of speech even in transition regions. The vectors of the excitation codebook are adapted to a training speech sequence by applying an iterative algorithm. To obtain a high coding accuracy, the analysis-by-synthesis error measure used during coding process is also used in the codebook design procedure. Due to the simultaneous occurance of quantized amplitude vector and quantized gain in the error measure, both codebooks are optimized iteratively. The amplitude codebook vectors are designed as subvectors of a so-called base excitation sequence by shifting their offset. Comparative listening tests have shown that this method outperforms stochastic CELP in objective SNR as well as in subjective quality.

1 INTRODUCTION

CELP [1] is a successful technique to keep a good speech quality at rates down to 6 kbit/s. For a further improvement of the performance of CELP schemes an advanced excitation design is one main topic. A promising method is based on the use of a trained excitation codebook instead of a stochastic codebook [2]. Here we present a new algorithm for the training of excitation codebooks. A great advantage of the proposed method is that only the codebook design process has to be modified, so no increase in complexity compared to the coder scheme of stochastic CELP is caused. Furthermore, the method is consistent to the analysis-by-synthesis (ABS) optimization criteria used in the CELP coding process and in particular it needs no additional assumptions about the characteristics of speech.

2 ANALYSIS OF THE CODING ERROR IN CELP SCHEMES

CELP coders using a stochastic codebook for coding the excitation are well suited for unvoiced speech segments. With long-term prediction, they also obtain a relatively high accuracy in stationary parts of voiced speech. Problems arise in transition regions where the memory of the long-term predictor isn't already adapted to the changed signal characteristics. The stochastic excitation on the other hand can't describe the residual in these complex parts of speech signals adequately and a poor quality results [3]. As shown in Figure 1, CELP schemes with a trained excitation codebook are able to reproduce more complex excitation shapes. This is due to the fact that an adaptation to different local characteristics of speech becomes possible.

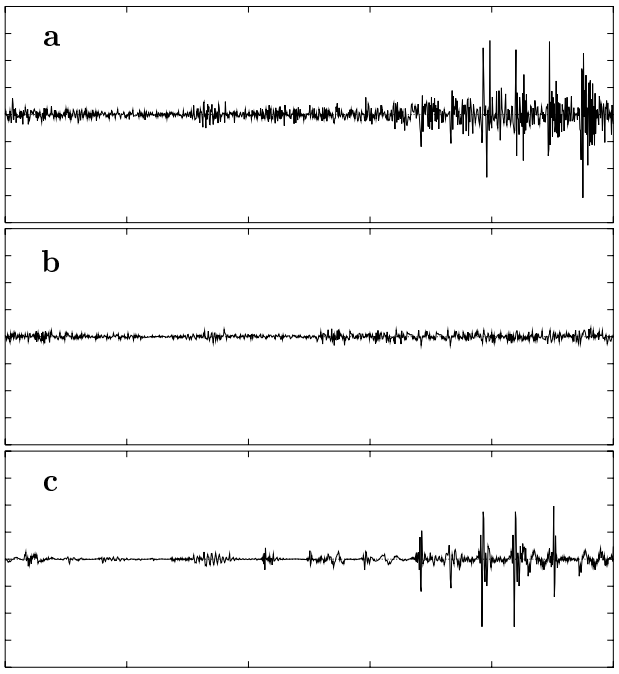

Figure 1: Residual of speech (a) and coded excitations of CELP schemes with stochastic excitation codebook (b) and with trained excitation codebook (c).

3 DESIGN OF TRAINED EXCITATION CODEBOOKS

In the proposed codebook design method, the excitation vectors are adapted to a training speech sequence by applying an iterative codebook optimization scheme which is an extension of the LBG algorithm [4]. According to LBG optimization two steps are performed iteratively. In the first step the training data is coded. In the second step the centroids of all data vectors which are assigned to the same codebook vector are calculated. For a minimum coding error the error measure in the optimization process has to be equal to the one used during coding. Therefore, the training data is coded using ABS with the psychoacoustically motivated weighting of the coding error and the centroids of the assignment regions are calculated consistently to this error measure as shown in the following.

For a given set of N_s-dimensional training data vectors $\{s_i(n)\}$ the accumulated coding error ε_j for all data vectors belonging to the assignment region P_j of a codebook vector $c_j(n)$ is given by

$$\varepsilon_j = \sum_{\substack{i \\ s_i(n)\in P_j}} \frac{\sum_{n=0}^{N_s-1} \left(s_i^W(n) - g_i \cdot c_j(n) * h_i^W(n)\right)^2}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (1)$$

where $s_i^W(n)$ is the ith data vector filtered by the weighting filter and g_i is the optimum value of the quantized gain relative to $s_i^W(n)$. $h_i^W(n)$ describes the impulse response of the cascaded synthesis filter and weighting filter. The denominator $\sum_{n=0}^{N_s-1} {s_i^W}^2(n)$ which usually doesn't occur in the ABS optimization criteria, keeps the signal to noise ratio (SNR) invariant from the signal energy. The consistency to the ABS optimization criteria isn't affected by this. Minimizing ε_j leads to a set of linear equations

$$\underline{\underline{R}}_j \cdot \underline{c}_j = \underline{r}_j \quad (2)$$

with

$$R_j(l,m) := \sum_{\substack{i \\ s_i(n)\in P_j}} \frac{\sum_{n=\max(l,m)}^{N_s-1} {g_i}^2 \cdot h_i^W(n-l) \cdot h_i^W(n-m)}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (3)$$

$$r_j(l) := \sum_{\substack{i \\ s_i(n)\in P_j}} \frac{\sum_{n=l}^{N_s-1} g_i \cdot s_i^W(n) \cdot h_i^W(n-l)}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (4)$$

for the new codebook vector $\underline{c}_j$. It can be solved by Cholesky's decomposition.

Because the quantized gain occurs simultaneously with the quantized amplitude vector in the error measure, a distinct improvement of coding accuracy can't be expected while both codebooks are optimized separately. Thus, in the proposed training algorithm each iteration of the amplitude codebook is followed by an iteration where the gain codebook is optimized. A diagram of this algorithm is shown in Figure 2.

To get a new quantized gain value g_k, the accumulated coding error

$$\varepsilon_k = \sum_{\substack{i \\ s_i(n)\in P_k}} \frac{\sum_{n=0}^{N_s-1} \left(s_i^W(n) - g_k \cdot c_i(n) * h_i^W(n)\right)^2}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (5)$$

for the assignment region P_k of g_k has to be minimized. According to this, g_k is given by

$$g_k = \frac{\sum_{\substack{i \\ s_i(n)\in P_k}} \Phi_{c_i * h_i^W, s_i^W} / \Phi_{s_i^W, s_i^W}}{\sum_{\substack{i \\ s_i(n)\in P_k}} \Phi_{c_i * h_i^W, c_i * h_i^W} / \Phi_{s_i^W, s_i^W}} \quad (6)$$

with

$$\Phi_{x,y} := \sum_{n=0}^{N_s-1} x(n) \cdot y(n). \quad (7)$$

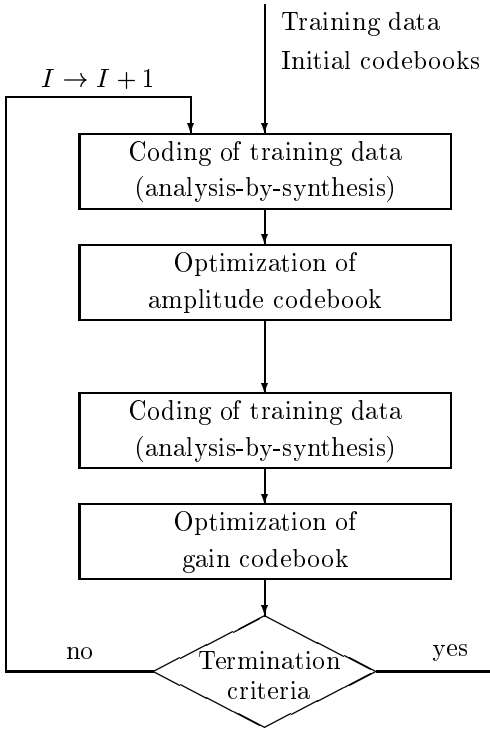

Figure 2: Flow-diagram of the algorithm for excitation codebook training.

4 CODEBOOK DESIGN WITH VARIABLE OFFSET

A further extension of the proposed algorithm takes into account that for a fixed frame length, characteristic waveforms in the residual occur with different offset to the frame boundaries. To synchronize the excitation shape with the frame boundaries, the codebook vectors are designed as parts of so-called base excitation sequences by shifting their offsets as it is shown in Figure 3. In this way each possible offset value is realized in the codebook. Then, each codebook vector is uniquely defined by the index j' of the base excitation sequence it belongs to and by its offset φ. The center of a base excitation sequence appears in the codebook with each possible offset to the frame boundaries. Whereas, the initial and final amplitudes of a base excitation sequence describe only the neighborhood of this particular central part.

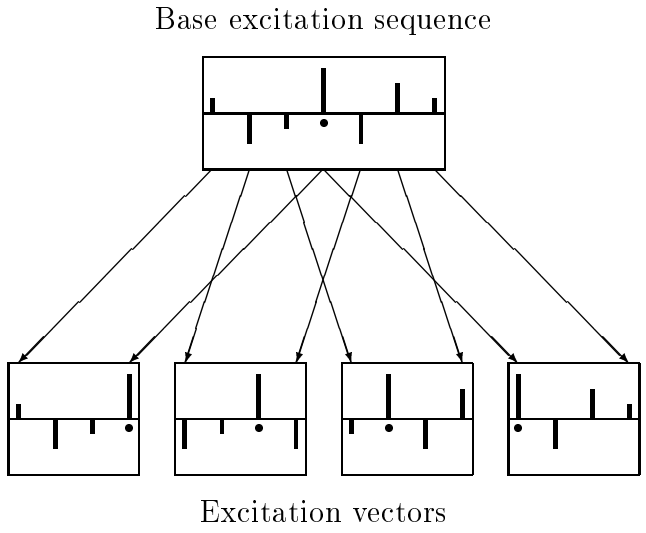

Figure 3: Codebook design with variable offset for the example of a single base excitation sequence consisting of $N_c = 7$ amplitudes and a frame length of $N_s = 4$ samples. The marked amplitude illustrates that each possible offset value is available in the codebook as excitation vector.

As a further advantage of this codebook design, the structure of the excitation codebook leads to a very low storage requirement without any loss in coding accuracy. Because each vector in the codebook differs from the previous one, only in one amplitude, for K vectors of dimension N_s, a base excitation sequence consisting of only $N_s + K - 1$ amplitudes has to be stored. E.g. for $N_s = 48$ and $K = 1024$ only 1071 amplitudes are neccessary instead of 49152.

In codebook training, data vectors with similar shape but with different offsets are assigned to codebook vectors which belong to the same base excitation sequence. By shifting, the optimum offset of each data vector is considered. Thus, it is possible to concentrate the adaptation of an excitation sequence to the different local characteristics of speech.

The optimization of the amplitude codebook which was deduced in (1) – (4), has to be modified in such a way that each codebook vector $c_j(n)_{n=0,1,\ldots,N_s-1}$ is described as the part $\tilde{c}_{j'}(n+\varphi)_{n=0,1,\ldots,N_s-1}$ of the base excitation sequence $\tilde{c}_{j'}(n)_{n=0,1,\ldots,N_c-1}$ the vector belongs to. According to this the accumulated coding error $\varepsilon_{j'}$ for the assignment region $P_{j'}$ of a base exitation sequence $\tilde{c}_{j'}(n)$ is given by

$$\varepsilon_{j'} = \sum_{\substack{i \\ s_i(n)\in P_{j'}}} \frac{\sum_{n=0}^{N_s-1} \left(s_i^W(n) - g_i \cdot \tilde{c}_{j'}(n+\varphi_i) * h_i^W(n)\right)^2}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (8)$$

where φ_i is the optimum offset value relative to the ith training data vector. Because the vectors of the excitation codebook are designed by shifting their offsets, each codebook vector overlaps with the following one in all amplitudes except of one. Thus, all codebook vectors assigned to the same excitation sequence depend on each other and the set of linear equations (2) can't be solved for each vector separately anymore. Instead of this, a coupled set of linear equations

$$\underline{\underline{\tilde{R}}}_{j'} \cdot \underline{\tilde{c}}_{j'} = \underline{\tilde{r}}_{j'} \quad (9)$$

with

$$\tilde{R}_{j'}(l,m) := \sum_{\substack{i \\ s_i(n)\in P_{j'} \\ [\varphi_i,\varphi_i+N_s-1]\supset\{l,m\}}} \frac{\sum_{n=\max(l,m)-\varphi_i}^{N_s-1} {g_i}^2 \cdot h_i^W(n-l+\varphi_i) \cdot h_i^W(n-m+\varphi_i)}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (10)$$

$$\tilde{r}_{j'}(l) := \sum_{\substack{i \\ s_i(n)\in P_{j'} \\ [\varphi_i,\varphi_i+N_s-1]\ni l}} \frac{\sum_{n=l-\varphi_i}^{N_s-1} g_i \cdot s_i^W(n) \cdot h_i^W(n-l+\varphi_i)}{\sum_{n=0}^{N_s-1} {s_i^W}^2(n)} \quad (11)$$

has to be solved for the whole base excitation sequence.

5 SIMULATION OF THE TRAINING PROCEDURE

For training an excitation codebook consisting of a 10 bit amplitude and a 3 bit gain codebook we have used a training speech sequence of 200 s phonetically balanced German speech. With a frame length of 48 samples for coding the excitation 33340 data vectors result. It has been found that the performance of codebook training depends very much on the choice of the initial codebooks. We gained the best results by applying a sequence of gaussian random numbers as initial

amplitude codebook and the square roots of the mean signal power in randomly choosen training vectors as initial gain codebook. During each optimization of amplitude as well as of gain codebook the segmental signal to noise ratio for the weighted training speech (SNR_W) was determined. As can be seen in Figure 4, the SNR_W increases dramatically in the first 10 iterations. Whereas, the increase of the following $15-20$ iterations leads to a further relevant increase in coding accuracy. By running the training algorithm on a 130 MFLOP/s workstation, for 30 iterations about 8 hours of CPU time are required.

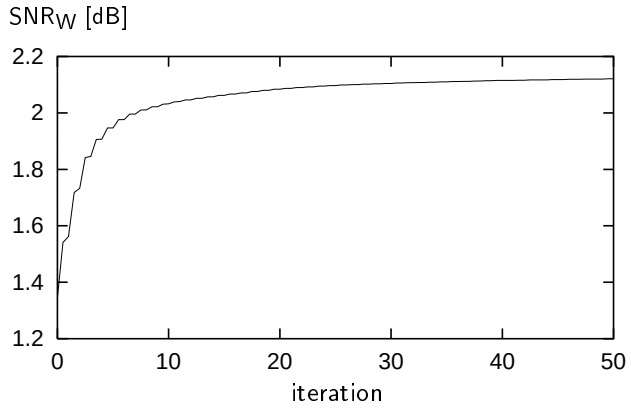

Figure 4: Increase of SNR_W [dB] during training procedure

6 PERFORMANCE ANALYSIS

The performance of the proposed codebook design method was examined for the coding of speech signals limited to telephone bandwidth. For comparison purposes a stochastic CELP scheme and a pulse-oriented CELP scheme, which is based on regularly spaced pulse grids (adaptive CELP [5, 6]), were configured, too. These systems differ from the trained CELP scheme only in the excitation. As it is shown in Table 1, a trained excitation codebook leads to a high increase in the SNR compared with stochastic CELP. The SNR is even higher than that of the pulse oriented CELP scheme. In comparative listening tests the CELP scheme with trained excitation also performed significantly better than the stochastic CELP scheme. The subjective intelligibility of the coded speech is found as high as that of the pulse oriented CELP scheme but the trained excitation design offers a more natural sounding speech. While the SNR for the CELP scheme with trained codebook increases only a few by using the codebook design with variable offset, comparative listening tests have shown a further distinct increase in speech quality.

Table 1: Comparison of SNR for different codebook designs for the coding of telephone bandwidth speech at a data rate of 4.8 kbit/s.

codebook design	SNR [dB]	seg. SNR [dB]
stochastic	5.946	4.455
pulse-oriented	7.137	5.229
trained:		
- fixed offset	7.396	5.251
- variable offset	7.810	5.421

7 CONCLUSIONS AND PERSPECTIVES

The proposed algorithm for the training of CELP excitation codebooks permits an adaptation of the excitation to the waveforms occuring in the residual of speech. By designing the excitation vectors from a base excitation sequence also the storage requirements for the excitation codebook can be drastically reduced. Comparative listening tests have shown that the proposed method outperforms stochastic CELP in terms of subjective inteligibility as well as in terms of a more natural sounding speech.

In future investigations, CELP schemes with trained excitation codebook are applied to wideband coding of speech and music signals.

References

[1] Atal, B.S., Schroeder, M.R., "Code Excited Linear Prediction (CELP)", Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Tampa (Florida), 1985, pp. 937-940.

[2] Chen, J.H., Gersho, A., "Vector Adaptive Predictive Coding of Speech at 9.6 kb/s", Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Tokyo, 1986, pp. 1693-1696.

[3] Atal, B.S., Caspers, B.E., "Beyond Multipulse and CELP towards High Quality Speech at 4 kb/s", in "Advances in Speech Coding", ed. Atal, B.S., Cuperman, V., Gersho, A., Kluwer Academic Puplishers 1990, pp. 191-201.

[4] Linde, Y., Buzo, A., Gray, R.M., "An Algorithm for Vector Quantizer Design", IEEE Trans. on Comm., vol. 28, pp. 84-95.

[5] Kipper, U., Reininger, H., Wolf, D., " Improved CELP Coding Using Adaptive Excitation Codebooks", Proc. ICASSP 91, Int. Conf. on Acoustics, Speech and Signal Processing, Toronto (Canada) 1991, vol. I, pp. 237-240.

[6] Kipper, U., Reininger, H., Wolf, D., " Low Bit Rate Speech Coding Using CELP with Adaptive Excitation Codebook", EUROSPEECH 91, 2nd European Conference on Speech Communication and Technology, Genua (Italy) 1991, vol. III, pp. 1353-1356.