Very Low Complexity CELP Speech Coding with Adaptive Phase Codebooks

Christian G. Gerlach

Institute of Communication Systems and Data Processing, Aachen University of Technology
52056 Aachen, Germany, fax: +49.241.8888186, e-mail: gerlach@ind.rwth-aachen.de

ABSTRACT

In analysis-by-synthesis speech coders the computational complexity of the search for an optimum innovation is still high though transformations were proposed to decrease the complexity e. g. [1]. This limits practical codebook sizes and vector dimensions (block lengths). In [2] a transform coding interpretation of analysis-by-synthesis was stated. Further it has been shown that unity magnitude codebooks in the frequency domain with the correct phase information allow near original speech quality. So unity magnitude codebooks with structured phases were proposed, that permit a computational complexity which increases only proportional to the bit rate and not to the codebook size. But in this simple approach slight perceptual degradations remained. In continuation to these ideas, an adaptive product codebook is proposed in this contribution, which allows more natural phase contours to eliminate the degradations, while keeping the complexity extreme low. Essential for the new concept, is the dynamic bit allocation to frequency intervals with variable bandwidth. Considerable improvements of both objective and subjective quality were achieved.

1. Introduction

In code-excited linear predictive (CELP) speech coding schemes using the analysis-by-synthesis technique [3] the perceptual quantization error at the output of a synthesis filter

$$E_p = \|\mathbf{z} - \gamma_q \mathbf{H}\mathbf{c}\|^2 \,, \; \gamma_q > 0, \; \mathbf{c} \in \mathrm{CB}, \; \gamma_q \in \mathrm{QT} \qquad (1)$$

has to be minimized by selecting a codevector $\mathbf{c} \in \mathbf{R}^L$ from a codebook CB and a scale factor γ_q from a quantization table QT. The target vector $\mathbf{z}$ usually consists of the weighted original signal after subtraction of the weighted contribution by the adaptive codebook and the weighted ringing of the synthesis filter [4]. $\mathbf{H}$ is the widely used filtering matrix consisting of shifted versions of the synthesis filter impulse response $h(n)$.

Usually the whole codebook represented by b bits must be searched through. The computational complexity is thus always linearly depending on the codebook size 2^b which limits possible number of bits. This remains true although structured ternary or algebraic [5] or vector summed [6] codebooks were proposed. The desire for larger frame sizes, possibly in the case of wideband speech coding, or for higher bit rates in variable bit rate applications ultimately limits the usefulness of these techniques.

Working in the frequency domain e. g. allows separation of the error measure in disjunctive frequency intervals. The idea of the new concept presented here, is to use a general adaptive product codebook in the frequency domain for the phases of a unity magnitude excitation. It is composed by independent codebooks for each frequency interval, while the bit rate per interval is dynamically adjusted. By generalizing the structured phase codebooks as proposed in [2] perceptual drawbacks are eliminated by using more natural phase contours.

2. Frequency Domain Representation

Considering the improved error criterion [4] or a similar version according to the considerations in [7], $\gamma_q \mathbf{H}\mathbf{c}$ can be expressed as a convolution of the excitation $\gamma_q c(n)$ with the truncated impulse response $h(n)$ of length R and

$$e(n) = z(n) - \gamma_q h(n) * c(n) \qquad (2)$$

is the error sequence of length $L + R - 1$ whose energy must be minimized. Given the zero padded sequences from (2) in the frequency domain with a DFT length of $N \geq L + R - 1$ the error criterion corresponding to (1) reads

$$E_p = \frac{1}{N} \sum_{k=0}^{N-1} |Z(k) - \gamma_q H(k) C(k)|^2 \rightarrow \min . \qquad (3)$$

Now we can interpret the analysis-by-synthesis quantization. As elaborated in [2] a correspondence to transform coding can be given. Due to the decoupling property of the Fourier transform

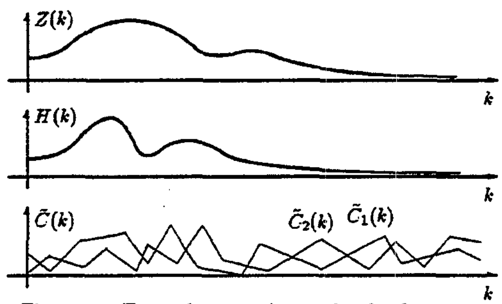

Figure 1: Example assuming real valued spectra

the spectral values $\tilde{C}(k) = \gamma_q C(k)$ play the role of normalized quantized transform coefficients and the analysis-by-synthesis technique can be interpreted as (distortion oriented) vector quantization of these coefficients by vectors $\tilde{C}_\nu(k) \in \mathrm{CB}_\mathcal{F}$.

3. Adaptable Product Code

We propose adaptable codebooks consisting of complex unity magnitude – precisely band pass – sequences (spectra) $C(k)$ according to

$$C(k) = |C(k)| e^{j\varphi_c(k)} = \begin{cases} 0 & \wedge\, k = 0 \\ 1 \cdot e^{j\varphi_c(k)} & \wedge\, k \in I_i \quad , i = 1 \ldots m \\ 0 & \wedge\, k = \frac{N}{2} \end{cases} . \qquad (4)$$

Since $C(k)$ corresponds to real valued time sequences, a definition for $0 \leq k \leq N/2$ is sufficient. The positive frequency axis is separated in m disjunctive intervals $I_i = [k_{i-1}, k_i - 1]$ with $i = 1,\ldots,m$. Denoting $Z(k)H^*(k) = B_{re}(k) + jB_{im}(k)$ with $B_{re}(k), B_{im}(k) \in \mathbf{R}$ we can state now, as an equivalent criterion for the squared quantization error

$$E_p' = 2 \sum_{k=1}^{\frac{N}{2}-1} \left(|Z(k)|^2 + \gamma_q^2 |H(k)|^2\right) \\ -2\gamma_q \sum_{k=1}^{\frac{N}{2}-1} (B_{re}(k)\cos\varphi_c(k) + B_{im}(k)\sin\varphi_c(k)) \to \min . \tag{5}$$

In order to minimize E_p' only the last sum in (5) has to be maximized by selecting the proper code sequence $\varphi_c(k)$. Consequently we suggest the phase sequences $\varphi_c(k)$ to be taken from a product codebook. Concatenating m phase sequences from phase codebooks $\varphi_i(k) \in \mathrm{CB}_i$ for these intervals, a product codebook [8, pp. 430] can be defined according to

$$\varphi_c(k) = \varphi_i(k) \text{ for } k \in I_i \text{ with } \varphi_i \in \mathrm{CB}_i \text{ and } i = 1,\ldots,m \quad . \tag{6}$$

According to the fixed range, the first and last interval are given to $I_1 = [1, k_1 - 1]$ and $I_m = \left[k_{m-1}, \frac{N}{2} - 1\right]$.

Due to the separability of the error in the frequency domain, the codeword selection can be performed independently for each interval I_i by maximizing

$$\sum_{k \in I_i} (B_{re}(k)\cos\varphi_i(k) + B_{im}(k)\sin\varphi_i(k)) \quad , \tag{7}$$

where the sequences $\cos\varphi_i(k)$ and $\sin\varphi_i(k)$, $k \in I_i$ belong to one of the entries of the codebook CB_i.

Now for example, each interval is represented by an equal number of bits (address bits of codebook CB_i) and the length of the frequency intervals is adapted according to $|H(k)|^2$ in the same way as derived in [2]. This means that short lengths are taken in regions where $|H(k)|^2$ is large (important regions) and wide lengths were $|H(k)|^2$ is small, thus providing a variable bit allocation (implicitly). This adaptation is essential to keep the high performance of analysis-by-synthesis while allowing low complexity.

4. Efficiency

If the number of bits for quantizing the sequence $c(n)$ is b, we have $\frac{b}{m}$ bits per interval or $2^{\frac{b}{m}}$ codevectors. It can be shown that the total number of multiplication and addition operations for the complete codebook search excluding the fixed amount of precomputations related to e. g. FFT's is only

$$2^{\frac{b}{m}} \cdot 2(N/2 - 1) \quad . \tag{8}$$

This can be further cut by a factor of 2 by exploiting a symmetric structure of the codebook per interval. If m is increased with the number of bits b the complexity can still be kept in bounds. Especially for m being only 2 or 3 the complexity is drastically reduced compared to the full search of a conventional codebook.

5. Application

The proposed algorithm was applied to a CELP codec (8 kbit/s) with closed loop pitch determination (adaptive codebook). First results confirm the predicted computational savings. If one uses $m = 3$ intervals and compares to a classical configuration with 1024 codewords for $L = 40$ samples using the autocorrelation method, the total complexity reduction for the stochastic codebook search results in a factor of about 24 (!) Further, if the bit rate for the excitation is increased now, no significant increase of the complexity has to be expected.

Of particular importance in this contribution is the quality improvement. Compared to the results of using a structured phase codebook with piecewise constant phases as described in [2] considerable improvements of both objective and subjective criteria were achieved. So for example the SNR increased from 6.4 dB to 7.3 dB by nearly 1 dB. Further, while the solution in [2] produced a synthesized signal with a slight crimpy or ripply noise, audible with headphones, this is eliminated by the proposed solution. This solution is now comparable in quality to using an unstructured unity magnitude codebook in a full search. Audio demonstrations can be given at the workshop. The complexity instead is dramatically reduced.

6. Summary and Conclusions

For the fixed codebook search in CELP algorithms we propose a general adaptive product codebook for the phases of a unity magnitude excitation. By dynamic bit allocation, the performance of analysis-by-synthesis is nearly kept, though a drastically reduced complexity in a strictly optimal determination scheme is achieved. This complexity no longer increases proportional to the codebook size, which is a prerequisite for using higher number of bits either due to larger frame sizes or higher bit rates for the excitation. Compared to previous results in [2] the perceptual drawbacks can be eliminated by generalizing the structured phase codebooks.

In contrast to other approaches a quantization method is derived which is more balanced between speech quality and computational complexity. This solution is achieved by combining three components: a perceptual based excitation model, gained insights in analysis-by-synthesis and techniques of complexity reduction. The derived general method is applicable to various linear predictive coding schemes.

References

[1] I. M. Trancoso and B. S. Atal, "Efficient search procedures for selecting the optimum innovation in stochastic coders," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 385–396, Mar. 1990.

[2] C. G. Gerlach, "CELP speech coding with almost no codebook search," in *Proc. of ICASSP-94*, (Adelaide, South Australia, Australia), pp. II 109–112, IEEE, Apr. 1994.

[3] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Transactions on Communications*, vol. 30, pp. 600–614, Apr. 1982.

[4] W. Kleijn, D. Krasinski, and R. Ketchum, "Improved speech quality and efficient vector quantization in SELP," in *Proc. of ICASSP-88*, (New York), pp. S4.4, 155–158, IEEE, 1988.

[5] J.-P. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in *Proc. of ICASSP-87*, pp. 1957–1960, IEEE, Apr. 1987.

[6] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8 kbs," in *Proc. of ICASSP-90*, Apr. 1990.

[7] C. G. Gerlach, "Analysis-by-synthesis speech coding with extended approximation interval," in *Proc. of IEEE Workshop on Speech Coding for Telecommunications*, (Sainte Adèle, Québec, Canada), pp. 33–34, IEEE, Oct. 1993.

[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Dordrecht: Kluwer Academic Publishers, 1992.