ANALYSIS-BY-SYNTHESIS SPEECH CODING
WITH EXTENDED APPROXIMATION INTERVAL

Christian G. Gerlach

Institute for Communication Systems and Data Processing, Aachen University of Technology
Templergraben 55, 52056 Aachen, Germany, phone +49.241.806979, fax +49.241.806985

ABSTRACT

In analysis-by-synthesis coders the problem of approximating the original signal by the synthesized signal is solved over a limited time interval only. In this contribution a systematic investigation of possible improvements by using extended or even unlimited intervals is presented.

1. INTRODUCTION

In linear predictive coding using analysis-by-synthesis techniques [1, 2] the quantization error at the output of a synthesis filter $\frac{1}{P(z)} \cdot \frac{1}{A(z)}$ shall be minimized as depicted in Figure 1. The signal is sequentially processed in block mode. For each
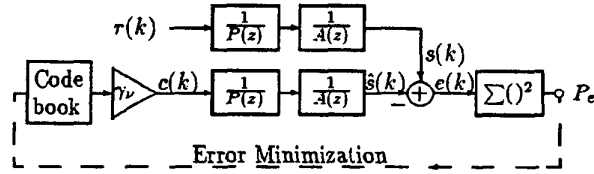

Figure 1: Basic structure of analysis-by-synthesis coders

whole block or interval I_ν consisting of L samples the quantized excitation sequence $c_\nu(k)$ is determined, so as to minimize the perceptual quantization error P_e. In the state-of-the-art schemes the error is computed just over the same time interval I_ν (analysis interval), ignoring the influence of the quantization on future intervals $I_{\nu+1}, I_{\nu+2}$ etc. In [3] this influence is taken into consideration to a certain extent only as the influence of excitation $c_\nu(k)$ on one future frame $I_{\nu+1}$ is considered, but by using an error P_e independent of $c_{\nu+1}(k)$.

In this paper a joint optimization technique is proposed, taking into account the contribution of the excitation sequences $c_\nu(k)$, $c_{\nu+1}(k)$, $\cdots$ to the expected distortion computed over all future frames. A theoretical solution for the described problem is derived and applied to a fixed grid Regular Pulse Excitation codec (RPE, e. g. [2]) and a Code Excited Linear Predictive codec (CELP, [1]). With this new systematic analysis, the influence of analysis time limitation in the traditional schemes can be evaluated.

2. SEQUENTIAL QUANTIZATION

The synthesis filter cascade in interval I_ν is described by its impulse responses $h_\nu(k)$ which can be truncated to $R \leq L$ samples. At first we consider the case that an FIR synthesis filter is used or that the coefficients of an IIR filter did not change from

$I_{\nu-1}$ to I_ν i. e. $h_{\nu-1} = h_\nu$. The synthesized signal in frame I_ν is denoted by $\hat{s}_\nu(k)$ and the input by $c_\nu(k)$. Then the vector $\hat{s}_\nu = (\hat{s}_\nu(1), \ldots, \hat{s}_\nu(L))^T \in \mathbf{R}^L$ is given by

$$\left(\begin{array}{cccccc} 0 \cdots & 0\, h_\nu(R\text{-}1) \cdots \\ \vdots & \ddots & \ddots \\ \vdots & & 0 \\ \vdots \\ 0 \cdots \cdots \cdots & \cdots \end{array} \middle| \begin{array}{cccccc} h_\nu(0) & 0 & \cdots & \cdots & 0 \\ \vdots & h_\nu(0) & 0 & & \vdots \\ h_\nu(R\text{-}1) & & \ddots & \ddots & \vdots \\ & \ddots & \ddots & & \ddots & 0 \\ \cdots & 0 & h_\nu(R\text{-}1) & \cdots & h_\nu(0) \end{array} \right) \left(\begin{array}{c} c_{\nu-1}(1) \\ c_{\nu-1}(2) \\ \vdots \\ c_{\nu-1}(L) \\ c_\nu(1) \\ c_\nu(2) \\ \vdots \\ c_\nu(L) \end{array} \right)$$

$$\underbrace{\hphantom{XXXXXXXX}}_{\mathbf{H}_\nu^1} \underbrace{\hphantom{XXXXXXXXXXX}}_{\mathbf{H}_\nu^0}$$

i. e. $\qquad \hat{s}_\nu = \mathbf{H}_\nu^1 \mathbf{c}_{\nu-1} + \mathbf{H}_\nu^0 \mathbf{c}_\nu .$ (1)

In the general case the matrices need not consist of time invariant shifted versions of the same impulse response, but all effects of time variation of the synthesis filter can be covered by proper specification of $\mathbf{H}_\nu^1$ and $\mathbf{H}_\nu^0$. The target vector $\mathbf{s}_\nu$ is equivalently substituted by $\mathbf{s}_\nu = \mathbf{H}_\nu^1 \mathbf{r}_{\nu-1} + \mathbf{H}_\nu^0 \mathbf{r}_\nu$ with $r_\nu(k)$ denoting the prediction residual sequence.

Exploiting the often existing delay of LPC-coders, a general premise shall be that for frame I_ν all target signals represented by $\mathbf{r}_i$ ($i \leq \nu+1$) and quantized excitation vectors $\mathbf{c}_i$ ($i \leq \nu-1$) are known.

The speech signal begins in frame I_1 thus $\mathbf{r}_0 = \mathbf{c}_0 = 0$. While M may be arbitrary large we obtain P_e summed over M frames as

$$P_e = \sum_{i=1}^{M} D_i^2 = \sum_{i=1}^{M} \|\mathbf{s}_i - \hat{\mathbf{s}}_i\|^2 = \sum_{i=1}^{M} \left\| \mathbf{H}_i^1 (\mathbf{r}_{i-1} - \mathbf{c}_{i-1}) + \mathbf{H}_i^0 (\mathbf{r}_i - \mathbf{c}_i) \right\|^2 .$$

(2)

3. MATHEMATICAL SOLUTION

Since the signal is only described statistically, the expected value of P_e shall be minimized by choosing the optimal excitation signal $\mathbf{c}_\nu$ in each frame. I. e. the task is to find a set of rules

$$\mathbf{c}_\nu = p_\nu(\mathbf{r}_{\nu-1} - \mathbf{c}_{\nu-1}, \mathbf{r}_\nu, \mathbf{r}_{\nu+1}) \quad \text{for} \quad \nu = 1, \ldots, M \quad (3)$$

that minimizes the expected value $E\{P_e\}$. The function p_ν has only three arguments due to the limited length of the impulse

response. P_e is a function of M random variables r_i and M functions p_i

$$E\{P_e\} = E\{P_e\{r_1,\ldots,r_\nu, p_1,\ldots,p_M\}\} = f(p_1,\ldots,p_M),$$
(4)

since the expectation value represents averaging over $r_1,\ldots,r_M$. The described problem is of the class treated by the calculus of variations. Following this theory e. g. [4, pp. 147] the existence of a solution is generally not guaranteed, therefore usually necessary conditions are imposed first.

A set of solution functions shall be $\varphi_1,\ldots,\varphi_M$ with $c_\nu^{opt} = \varphi_\nu(r_{\nu-1}-c_{\nu-1},r_\nu r_{\nu+1})$. By using the minimum conditions for φ_ν and $\varphi_{\nu+1}$ some inequalities can be exploited as shown in [5] to derive a necessary condition for a solution φ_ν. The essential result is stated below:

If $A(c_\nu,c_\nu^{opt})$ is defined as

$$A(c_\nu,c_\nu^{opt}) = \left\| \mathbf{H}_\nu^1(r_{\nu-1}-c_{\nu-1}) + \mathbf{H}_\nu^0(r_\nu-c_\nu) \right\|^2 +$$
$$\left\| \mathbf{H}_{\nu+1}^1(r_\nu-c_\nu) + \mathbf{H}_{\nu+1}^0\left(r_{\nu+1} - \right.\right.$$
$$\left.\left. E\{\varphi_{\nu+1}(r_\nu-c_\nu^{opt}, r_{\nu+1}, r_{\nu+2})|r_\nu-c_\nu^{opt}, r_{\nu+1}\} \right) \right\|^2$$
(5)

with the given function $\varphi_{\nu+1}$, the following relation

$$A(c_\nu,c_\nu^{opt}) \geq A(c_\nu^{opt},c_\nu^{opt})$$
(6)

holds for all $c_\nu \in \mathbf{R}^L$. This is an implicit criterion for c_ν^{opt}. Inserting a test vector $\tilde{c}_\nu^{opt}$ leads to $A(c_\nu, \tilde{c}_\nu^{opt})$. If the minimum of this function is taken for $c_\nu = \tilde{c}_\nu^{opt}$, then the necessary condition for $\tilde{c}_\nu^{opt}$ being an optimal value is fulfilled. With this criterion the problem can generally be solved; at least theoretically. Beginning with $\varphi_M(\cdots)$ and ending with $\varphi_1(\cdots)$ all functions can be obtained recursively. Practically, approximations were found that allow a forward directed proceeding.

4. APPLICATION

By using condition (6) for the fixed grid RPE-approach an analytical solution was derived [5], which was applied to a speech codec based on the GSM-codec and provided an improvement of 0.2 dB averaged SNR but 3 dB for sine signals.

In the case of vector quantization when c_ν is a codevector chosen from a codebook, analytical solutions can not be expected. However equation (5) and (6) can be interpreted in an instructive way: If the expected value of the future optimum excitation vector $E\{c_{\nu+1}^{opt}|\cdots\}$ were given, the optimum c_ν^{opt} could be computed by minimizing the special squared error over two frames as noted in (5).

Using estimated values $\hat{c}_{\nu+1}^{opt}$ for $E\{c_{\nu+1}^{opt}|\cdots\}$ provides the basis to develop new improved sequential quantization methods as an approximation of the theoretic solution. One proposed scheme is derived as follows: $\hat{c}_{\nu+1}^{opt}$ itself is determined by condition (5) and (6) for index $\nu+1$. In parallel, in this condition $E\{c_{\nu+2}^{opt}|\cdots\}$ is used. If the vector $c_{\nu+2}^{opt}$ is now approximated by $r_{\nu+2}$ the equation $r_{\nu+2} - E\{c_{\nu+2}^{opt}|\cdots\} = 0$ holds. Consequently, the condition reduces to minimizing

$$\left\| \mathbf{H}_{\nu+1}^1(r_\nu-c_\nu) + \mathbf{H}_{\nu+1}^0(r_{\nu+1} - \hat{c}_{\nu+1}^{opt}) \right\|^2 + \left\| \mathbf{H}_{\nu+2}^1(r_{\nu+1} - \hat{c}_{\nu+1}^{opt}) \right\|^2$$
(7)

which leads to $\hat{c}_{\nu+1}^{opt}$. If this $\hat{c}_{\nu+1}^{opt}$ is used for determination of $\hat{c}_\nu^{opt}$ from (5) and (6), an approximate solution for c_ν^{opt} can be found. This must be done by iteration since in (7) c_ν is already included. The iteration starts with $r_\nu-c_\nu = 0$ in (7). Then (5) is used to find an iteration value of $\hat{c}_\nu^{opt}$ which in turn can be used in (7) to improve the iteration value of $\hat{c}_{\nu+1}^{opt}$. If $\hat{c}_\nu^{opt}$ does not change any more (fixed point), the iteration stops and $\hat{c}_\nu^{opt}$ is taken as the quantized excitation sequence for frame I_ν.

This method was implemented for simple 8 kHz CELP-codecs without noise weighting nor long term predictor [6]. In 99% of all blocks one or two iterations were sufficient to reach a fixed point. So the number of iterations was always limited to two.

To judge the gained improvements the absolute maximum threshold for quantization with the given codebook was determined. It is computed by an exhaustive search over all codeword combinations for an entire file which implies that the whole signal must be known before the quantization starts.

We made simulations for block lengths $L = 20$, $L = 40$ and different codebook sizes. The following values for this simple codec operating at approximately 5 dB SNR will exemplify the general trend. The difference in SNR between the search with time limited analysis interval and the noncausal absolute threshold was only 0.39 dB. From this margin the proposed algorithm achieves already 0.34 dB.

Using the proposed algorithm will raise the SNR without changing an existing decoder or the channel structure. Since it promises more improvement for smaller block lengths, it was implemented for the CCITT standard G.728 LD-CELP codec and increased the SNR by 0.6 dB on the average and 1 dB in good cases starting from approximately 17 dB. The introduced delay was just 0.625 ms. Therefore it is a feasible addition compatible with existing systems.

5. SUMMARY AND CONCLUSIONS

A systematic approach for judging the performance of the time limited analysis-by-synthesis is given.

The conventional algorithm is not very far away from the noncausal absolute optimum. The new proposed optimal sequential quantization practically comes up to the absolute optimum while exploiting solely one future frame.

REFERENCES

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

[2] P. Kroon, E. F. Deprettere, and R. J. Sluijter, "Regular-pulse excitation—a novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 1054–1063, Oct. 1986.

[3] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum, "Analysis and improvement of the vector quantization in selp," in *Proc. of EUSIPCO-88*, pp. 1043–1046, Sept. 1988.

[4] R. Courant and D. Hilbert, *Methoden der mathematischen Physik I*. Berlin, Heidelberg, New York: Springer-Verlag, 3 ed., 1968.

[5] C. G. Gerlach, "Optimality of sequential quantization in analysis-by-synthesis speech codecs," (Berlin), Sept. 1993. presented at Eurospeech 93.

[6] A. Busboom, July 1993. "Sprachcodierung mit sequentieller Analyse durch Synthese", diploma thesis at Institut für Nachrichtengeräte der RWTH Aachen.