FLEXIBLE QUANTIZATION OF AUDIO AND SPEECH BASED ON THE AUTOREGRESSIVE MODEL

Alexey Ozerov and W. Bastiaan Kleijn

ACCESS Linnaeus Center, Electrical Engineering
KTH - Royal Institute of Technology
Stockholm, Sweden
{alexey.ozerov,bastiaan.kleijn}@ee.kth.se

ABSTRACT

We describe a coding scheme based on audio and speech quantization with an adaptive quantizer derived from the autoregressive model under high-rate assumptions. The main advantage of this scheme compared to state-of-the-art training-based coders is its flexibility. The scheme can adapt in real time to any particular rate and has a computational complexity independent of the rate. Experiments indicate that, compared with a non-scalable conventional fixed-rate code-excited linear predictive (CELP) coding scheme, our real time scalable coder with scalar quantization performs at least as well in the constrained entropy case, and has nearly identical performance for the constrained resolution case.

1. INTRODUCTION

Transmission networks are becoming increasingly heterogeneous. The new network environment requires a new generation of flexible coders that are able to adapt in real time to continuously changing network conditions, and particulary to varying transmission rate. Most existing coders (e.g., CELP coders) require rate-dependent off-line training, and, therefore, do not support such flexibility.

The modern heterogeneous environment is not well served by the ubiquitous CELP algorithm, since:

- *CELP is not adaptable to any rate from a continuum of rates.* For every particular rate a new codebook must be trained. Thus it is possible to use CELP for a predefined set of rates, as for example in AMR-WB [1], but not for a continuum of rates.
- *Computational complexity grows exponentially with rate*, since it grows linearly with codebook size.
- *Storage requirements grow exponentially with rate,* since the pre-trained codebooks must be stored.
- *Quantization cell shapes are not locally optimal in signal domain.* In CELP, a fixed codebook in the excitation domain is mapped by a non-unitary transform (obtained from linear predictive coding (LPC) coefficients) to the signal domain for every quantization frame. Since a non-unitary transform does not preserve the squared error, the cell shapes can be only optimal in excitation domain, not in signal domain. Even if the codebook is trained minimizing the squared error in the signal domain, the cell shapes can be optimal only on average, but not for any local signal statistics.

In this paper we propose a flexible coding scheme that addresses all above-mentioned CELP shortcomings, while ensuring performance comparable with training-based CELP algorithm. In order to assure the rate-adaptability property of such a system, it should be able to compute in real time quantizers appropriate for any specified rate. The use of probabilistic source models, combined with high-rate theory approximations [2], facilitates the practical implementation of such computable quantizers. This approach allows a computational complexity independent of the rate [3]. It has low storage requirements independent of the rate (no quantization tables need to be stored), and facilitates optimal variable-rate quantization (constrained entropy quantization [4]). Moreover, it uses quantizers with optimal (for scalar quantization and under high-rate theory assumptions) cell shapes for the local model-defined signal statistics.

For the probabilistic source model we use the autoregressive (AR) model, which is commonly used to model the speech signal [5]. For each fixed-length signal block (frame) our approach consists of two steps. The first one is the AR model estimation and quantization of its parameters using some pre-trained Gaussian mixture model (GMM) of the line spectral frequencies (LSF) [3]. The second one is the signal frame quantization using the quantized AR model. Both model and signal are quantized with probabilistic model-based computable quantizers. Moreover, we study both constrained resolution (CR) (constant rate) and constrained entropy (CE) (variable rate) quantization scenarios.

Our approach is related to [6, 7, 8]. Kim and Kleijn [6] demonstrated that it is efficient to use a Gaussian signal model to encode a speech segment directly, showing both theoretical and practical advantages over the ubiquitous CELP coding scheme. Samuelsson [7] applied the GMM model based quantization approach introduced by [3] directly to the coding of the audio signal. We go beyond these contributions: we use a single Gaussian model for each signal block and use a theoretically optimal distribution of rate between signal and model [9]. Li and Kleijn [8] recently presented a low delay coder using CE signal quantization with AR model adapted in the *backward* manner, while in this paper we study the *forward* AR model adaptation case.

The paper is organized as follows. The principles of Gaussian model-based quantization under high-rate assumptions are described in section 2. The architecture of our flexible coder is presented in section 3. The results and conclusions are given in sections 4 and 5 respectively.

This work was supported in part by the European Union under Grant FP6-2002-IST-C 020023-2 FlexCode

978-1-4244-2110-7/08/$25.00 ©2007 IEEE

2. GAUSSIAN MODEL-BASED QUANTIZATION UNDER HIGH-RATE ASSUMPTIONS

In this section we recall the principles of practical quantization schemes with computable quantizers based on a single Gaussian source model under high-rate theory assumptions. These computable quantizers based on scalar quantization in the mean-removed Karhunen-Loeve transform (KLT) domain are described for both CR and CE cases. Extensions of these methods to the GMM case can be found in [3] for CR case and in [4] for CE case.

We consider a K-dimensional source vector $s = [s_1, \ldots, s_K]^T$, which is a realization of a random Gaussian vector $S = [S_1, \ldots, S_K]^T$ with mean vector μ and covariance matrix Σ (i.e., $S \sim \mathcal{N}(\mu, \Sigma)$). Let $\Sigma = U \Lambda U^T$ be the eigenvalue decomposition of the covariance matrix, i.e., U is an orthogonal matrix ($U^T U = I$) and $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\}$ is a diagonal matrix of eigenvalues.

We consider a CR quantization of a source vector s using a fixed budget of B bits and mean squared error (MSE) distortion. The quantization procedure is based on the companded scalar quantizers in the mean-removed KLT domain, and it consists of the following steps [3]:

1. Remove mean, apply the KLT, and normalize for standard deviation $\sqrt{\lambda_k}$ (Eq. (1)).
2. Apply the optimal (for the Gaussian random variable) scalar compressor, independently for each dimension k (Eq. (2)), where $\phi(\cdot)$ is the cumulative distribution function for a Gaussian random variable with zero mean and unit variance.
3. Apply a scalar quantizer uniform on the interval $(0, 1)$ with $L_k = 2^{B_k}$ levels (Eq. (3)), where $[\cdot]$ denotes rounding and B_k is computed as in Eq. (4).
4. Compute codeword index (Eq. (5)), and transmit it to the decoder.
5. At the decoder side the quantized source vector $\hat{s}$ is reconstructed from $\hat{i}$ by applying the inverse transforms corresponding to equations (5), (2) and (1).

$$\begin{aligned} x &= \Lambda^{-1/2} U^T (s - \mu) & (1) \\ u_k &= \phi(x_k/\sqrt{3}) & (2) \\ \hat{d}_k &= ([d_k L_k - 0.5] + 0.5)/L_k & (3) \\ B_k &= B/K + 0.5 \log_2 \left(\lambda_k / \prod_{l=1}^{K} \lambda_l^{1/K} \right) & (4) \\ \hat{i} &= 1 + \sum_{k=1}^{K} (\hat{d}_k - 1) \prod_{l=1}^{k-1} L_k & (5) \end{aligned}$$

For the CE case with MSE distortion, uniform quantization is asymptotically optimal [2]. Here we use scalar uniform quantizers of fixed step size Δ in the mean-removed KLT domain, and the quantization procedure can be summarized as follows [4]:

1. Remove mean and apply the KLT (Eq. (6)).
2. Quantize each dimension with a uniform scalar quantizer having a constant step size Δ (Eqs. (7) and (8)).
3. Again for each dimension, encode quantization index $\hat{d}_k$ with a lossless entropy coder using the distribution $p_{\hat{D}_k}(\cdot)$ defined by signal model as in equation (9), where $p_{Y_k}(\cdot)$ is the probability density function (pdf) of Gaussian random variable Y_k with zero mean and variance λ_k (i.e., $Y_k \sim \mathcal{N}(0, \lambda_k)$). Note that y_k is a particular realization of this random variable. We use an arithmetic coder as an entropy coder, which gives an effective codeword length l_k (in bits) given by equation (10).
4. At the decoder the quantized source vector $\hat{s}$ is reconstructed from vector of indices $\hat{d} = [\hat{d}_1, \ldots, \hat{d}_K]$ by applying equation (8) and the inverse of the transform described by equation (6).

$$\begin{aligned} y &= U^T (s - \mu) & (6) \\ \hat{d}_k &= [y_k/\Delta] & (7) \\ \hat{y}_k &= \hat{d}_k \Delta & (8) \\ p_{\hat{D}_k}(\hat{d}_k) &= \int_{\hat{y}_k - \Delta/2}^{\hat{y}_k + \Delta/2} p_{Y_k}(y_k) dy_k & (9) \\ l_k &= -\log_2(p_{\hat{D}_k}(\hat{d}_k)) & (10) \end{aligned}$$

Note that both the described CR and CE quantization schemes can run for any particular rate without any re-training, and have computational complexities independent of the rate.

3. CODING ARCHITECTURE

In the proposed coding scheme the sampled input audio signal is segmented into long blocks (frames) of 20 ms, and consequently into small blocks (subframes) of fixed length of K samples. A set of p-order linear predictive coding (LPC) coefficients is estimated for every frame and interpolated (in the LSF domain) for subframes. An excitation variance is estimated for every subframe.

3.1. Signal model

Let $\{a_i\}_{i=1}^{p}$ and σ^2 be the LPC coeffitients and excitation variance estimated for subframe s. These parameters correspond to the following AR model:

$$\frac{\sigma}{A(z)} = \frac{\sigma}{1 + a_1 z^{-1} + \ldots + a_p z^{-p}}. \qquad (11)$$

For quantization of subframe s we use a two-step redundancy removal: (i) the *intra-frame redundancy* is removed by AR model-based KLT, (ii) the *inter-frame redundancy* is removed by AR model-based "ringing" (or zero-input response) subtraction. Altogether these two redundancy removal steps are translated by modeling subframe s by a Gaussian random vector S distributed as:

$$S \sim \mathcal{N}(r, \Sigma), \qquad (12)$$

where the mean vector r is the ringing from previous subframe computed using LPC filter $1/A(z)$, and the covariance matrix Σ is computed as:

$$\Sigma = \sigma^2 (A^T A)^{-1}, \qquad (13)$$

where A is a lower-triangular ($K \times K$) Toeplitz matrix with as first column $[1, a_1, \ldots, a_p, 0, \ldots, 0]^T$.

Since the model parameters must be known as well at the decoder side, they are also quantized and transmitted. Thus, instead of the covariance matrix Σ from (12) we use its quantized version $\hat{\Sigma}$ computed from quantized LPC coefficients $\{\hat{a}_i\}_{i=1}^{p}$ and quantized variance $\hat{\sigma}^2$. Instead of the ringing r from (12) we use the ringing $\hat{r}$, which is computed from previous quantized subframe using LPC filter $1/\hat{A}(z)$.

3.2. Signal normalization

In this paper we do not consider any advanced model of perception, and the subframes are quantized using the MSE distortion. However, in order to fit the segmental signal to noise ratio (SSNR) criterion which we use for evaluation, every subframe must be normalized by some gain $\hat{g}$ representing it's energy. Note that this normalization makes sense for CE quantization only, since it governs the number of bits spent for quantization of every particular subframe. In the CR case, it does not change anything, since the same number of bits is spent for each subframe. The gain $\hat{g}$ is estimated as $\hat{g} = \sqrt{E[S^T S]/K}$, which, given (12), can be expressed as:

$$\hat{g} = \sqrt{(\hat{r}^2 + \mathrm{tr}[\hat{\Sigma}])/K}, \qquad (14)$$

where $\mathrm{tr}[\hat{\Sigma}]$ denotes the trace of matrix $\hat{\Sigma}$. Note that this normalization is a simple model of perception. More advanced perceptual models, like for example a perceptual weighting filter used in the AMR-WB coder [1], can easily be integrated in our coding scheme (see, e.g., [8]).

3.3. Coder structure

The structure of the proposed coder is shown on figure 1. Each quantizer of the coder is based on a probabilistic model and can run in either CR or CE quantization mode (see section 2 or [3, 4]). The LPC are quantized using a GMM in the LSF domain. The variance σ^2 is quantized using a single Gaussian in the log-domain. The signal waveform (subframe) is quantized using a multivariate Gaussian distribution (12). Thus, all the quantizers are computable and can run at any rate.

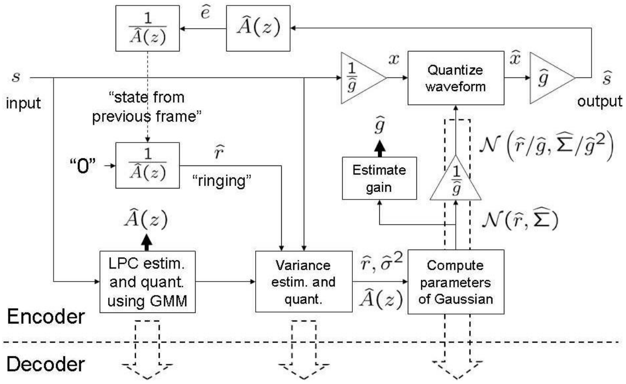

Fig. 1. Block diagram of the proposed encoder.

For each frame the LPC coefficients are estimated as described in the AMR-WB [1] coder specification, converted to the LSFs and quantized using some pre-trained GMM [3]. For subframes the quantized LSFs are interpolated as in AMR-WB [1] and converted back to LPCs. Thus, for each subframe s we obtain a set of LPC coefficients $\{\hat{a}_i\}_{i=1}^p$.

The "ringing" $\hat{r}$ from the previous subframe is computed by passing zero excitation through the LPC filter $1/\hat{A}(z)$. The state of this filter is updated by filtering through it the quantized excitation $\hat{e}$ from the previous subframe (see Fig. 1).

The AR model variance σ^2 is estimated for each subframe as:

$$\sigma^2 = (e^T e)/K, \qquad (15)$$

where e is the excitation computed by filtering the current subframe without ringing (i.e., $s - \hat{r}$) through the filter $\hat{A}(z)$ with zero initial state. This estimate is the maximum likelihood (ML) estimate of the AR model variance. The variance is quantized in log-domain using a single Gaussian distribution.

Given the ringing $\hat{r}$, quantized and interpolated LPC coefficients $\{\hat{a}_i\}_{i=1}^p$, and quantized variance $\hat{\sigma}^2$, the parameters of multivariate Gaussian distribution $\mathcal{N}(\hat{r}, \hat{\Sigma})$ are derived as described in section 3.1.

The normalization gain $\hat{g}$ is computed using (14). The normalized target vector $x = s/\hat{g}$ is quantized using the corresponding normalized multivariate Gaussian distribution $\mathcal{N}(\hat{r}/\hat{g}, \hat{\Sigma}/\hat{g}^2)$. The quantized subframe $\hat{s}$ is obtained from the quantized target vector $\hat{x}$ by multiplying it by the gain $\hat{g}$.

3.4. Rate distribution between signal and model

Since both the signal s and the AR model parameters are quantized, and the proposed system should adapt to any particular total rate, we must decide in real time for each rate, what is the optimal rate distribution between signal and model. In our recent study [9] we showed theoretically that under high-rate assumptions the optimal rate for model quantization is constant, i.e., it is independent on the total rate. This fact significantly simplifies the design of our flexible coding scheme. It means that when the total rate changes, only the rate used for signal must be adjusted, and the rate for model must be kept constant.

3.5. Ringing control

We have noted experimentally that for low rates (1-2 bits per sample) and for small subframe sizes ($K = 5$ - 10 samples) the present system can become unstable. This instability problem arises from the fact that there is a closed-loop between the ringing computation and estimation of the model variance (see Fig. 1). The conceptual problem is that the LPC coefficients $\{a_i\}_{i=1}^p$ and the variance σ^2 are quantized, thus constrained, but ringing $\hat{r}$, being a part of the model as well, is not constrained at all. Thus for some low rate, when there are not enough bits to quantize some subframe, the ringing for the next subframe can be quite different from the "true ringing" r (i.e., ringing computed from the non-quantized excitation), and then all estimations can diverge because of the above-mentioned closed-loop. We propose two different solutions allowing to avoid this instability problem for the CR and CE cases respectively.

For the CR case the solution consists of estimating the variance σ^2 based on the true ringing r computed from the non-quantized previous subframe s instead of the quantized $\hat{s}$. This breaks the above-mentioned closed-loop, thus solving the instability problem. This solution is satisfactory for the CR case, but not for the CE case, since it decreases the likelihood of the signal (i.e., $N(s; \hat{r}, \hat{\Sigma})$) provoking outliers, resulting in unacceptable bursts in the rate variation.

The solution for the CE case consists of introducing a modified distortion measure controlling the ringing variation during quantization of the signal. Assuming that the model order p is not greater than the subframe length K [1], the ringing for the next subframe $\hat{r}_{\mathrm{nxt}}$ is related to current quantized subframe $\hat{s}$ through

[1] In case when $p > K$, we just truncate the transform $B(\hat{a})$ so that it fits to the subframe length K.

some linear transform $B(\hat{a})$ (i.e., $\hat{r}_{\mathrm{nxt}} = B(\hat{a})\hat{s}$), which depends on LPC coefficients $\hat{a}$. The same is valid for true ringing r_{nxt}, i.e., $r_{\mathrm{nxt}} = B(\hat{a})s$. We can define the following weighted distortion measure:

$$d_w(s, \hat{s}) = [(1-\beta)\|s-\hat{s}\|^2 + \beta\|B(\hat{a})(s-\hat{s})\|^2]/\hat{g}^2, \quad (16)$$

where $\hat{g}$ is the normalization gain, and $\beta \in [0, 1)$ is a constant penalty factor that is small for low rates and zero for high rates (the zero value corresponds to the standard unweighted case). The distortion measure (16) can be represented as $d_w(s, \hat{s}) = (s - \hat{s})^T W(\hat{a}, \hat{g})(s - \hat{s})$, where $W(\hat{a}, \hat{g}) = [(1-\beta)^2 I + \beta^2 B(\hat{a})^T B(\hat{a})]/\hat{g}^2$ is a positive-definite *sensitivity matrix*. By computing the Cholesky decomposition of the sensitivity matrix one can obtain a linear transform H, such that for $v = Hs$ distortion $d_w(s, \hat{s})$ becomes a squared error (see, e.g., [10] for more details). Thus, CE quantization can be applied to v as described in section 2. This solution eliminates the instability problem in the CE case, and does not lead to an unacceptable rate variation.

3.6. Computational complexity and storage requirements

The computational complexity of the proposed encoder, excluding the complexity of the arithmetic coder, is summarized in table 1 with L being the length of a frame used for LPC estimation and M being number of components of GMM used for LPC quantization. The decoder's computational complexity is lower, since the decoder does not involve the estimation steps. We see once again that the computational complexity does not depend on the rate, and it is quite low, except the eigenvalue decomposition (EVD) computation, which is of the order of $O(K^3)$.

Table 1. Computational complexity of the proposed encoder.

Operation		every	Complexity
LPC coefficients estimation		L samples	$O(L^2)$
LPC coefficients quantization		L samples	$O(Mp^2)$
Variance estimation & quant.		K samples	$O(Kp)$
Signal subframe quantization	EVD	K samples	$O(K^3)$
	quantization	K samples	$O(K^2)$

The storage requirements are low. The only values that must be stored are the parameters of the GMM used for LPC quantization, which is of the order of $O(Mp^2)$ values.

4. RESULTS

4.1. Comparison with CELP

We first compare the proposed coding scheme for both CR and CE cases with a fixed-rate CELP coding scheme using a fixed excitation codebook trained optimizing the signal domain squared error criterion, as in [11]. For the comparison we used 10 sentences of narrowband speech from TIMIT database. For all compared schemes the LPC coefficients were not quantized, the subframe length was five samples ($K = 5$), and the total rate was 19.2 kbps, which is equivalent to 12 bits per subframe. The short subframe length was chosen since the reference CELP scheme computational complexity grows exponentially when K increases (given the constant rate in kbps).

The results in terms of SSNR together with optimal rate distributions between variance (or gain in case of CELP) and signal rates (experimentally adjusted for every scheme) are given in table 2. We see that, compared with a CELP with a codebook retrained for each tested rate, the performance of our scalable coder in terms of SSNR is at least as good in the CE case and almost as good in the CR case. It should be also mentioned that we use scalar quantizers in the mean-removed KLT domain, while CELP is based on vector quantization. By increasing K from 5 to 10 samples, and evaluating the proposed scheme for the same total rate of 19.2 kbps, we obtain 18.5 dB and 20.43 dB SSNR for CR and CE cases respectively. We are not able to perform practical simulations for CELP using the same setting because of the excessive computational load and memory requirements.

Table 2. Comparison of the proposed scheme with CELP.

	CR case	CE case	CELP
Variance rate (bits/subframe)	3	2.7	5
Signal rate (bits/subframe)	9	9.2	7
SSNR (dB)	16.07	17.96	17.82

4.2. Rate vs. distortion

In this section we present the rate-distortion equations for the studied model-based signal quantizers under high-rate assumptions. Since, as explained in section 3.4, the optimal rate spent for model quantization is constant whatever the total rate is, we present these rate-distortion equations only for the rate spent for signal transmission. Let x be quantized K-dimensional signal subframe drawn from a random Gaussian vector X. The distribution of X is defined by some model $\hat{\theta}(x) = \{\hat{r}(x), \hat{\Sigma}(x)\}$, which is assumed to be already estimated and quantized. Let $\hat{\Sigma}(x) = \hat{U}(x)\hat{\Lambda}(x)\hat{U}(x)^T$ be the eigenvalue decomposition of covariance matrix $\hat{\Sigma}(x)$.

In the CR case under high-rate assumptions the constant rate R_{CR} (in bits per subframe) is related to the average distortion $\bar{D}_{\mathrm{CR}}$ (per dimension) as [2]:

$$R_{\mathrm{CR}} = -\frac{K}{2}\log_2\left(\frac{\bar{D}_{\mathrm{CR}}}{C}\right) + \frac{K}{2}\log_2\left(\frac{3(2\pi)^{2/3}}{K}\right) + \frac{K}{2}\log_2 E\left[\prod_l \lambda_l(X)^{\frac{1}{K}} \sum_k \frac{N(y_k(X); 0, \lambda_k(X))^{-\frac{2}{3}}}{\lambda_k(X)^{\frac{1}{3}}}\right], \quad (17)$$

where $y(x) = \hat{U}(x)^T(x - \hat{r}(x))$ ($y(X) \sim \mathcal{N}(0, \hat{\Lambda}(x))$) is the vector x in the mean-removed KLT domain, $N(\cdot; 0, \lambda_k(x))$ is the pdf of its k-th component, $C = 1/12$ is the coefficient of quantization of a scalar quantizer, and the expectation operation $E[\cdot]$ means the averaging over all quantized subframes x.

In the CE case the average rate $\bar{R}_{\mathrm{CE}}$ (in bits per subframe) relates to the constant distortion D_{CE} (per dimension) as [12]:

$$\bar{R}_{\mathrm{CE}} = -\frac{K}{2}\log_2\left(\frac{D_{\mathrm{CE}}}{C}\right) - E[\log_2(p_{X|\hat{\Theta}}(X|\hat{\theta}(X)))], \quad (18)$$

where $p_{X|\hat{\Theta}}(\cdot|\hat{\theta}(x))$ is the pdf of X.

[2] Relation (17) is derived from the rate-distortion function for scalar quantizers from [12], using the rate distribution described by equation (4).

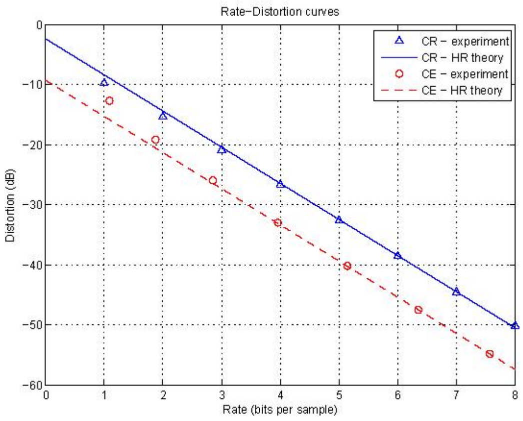

Fig. 2. Rate-Distortion curves.

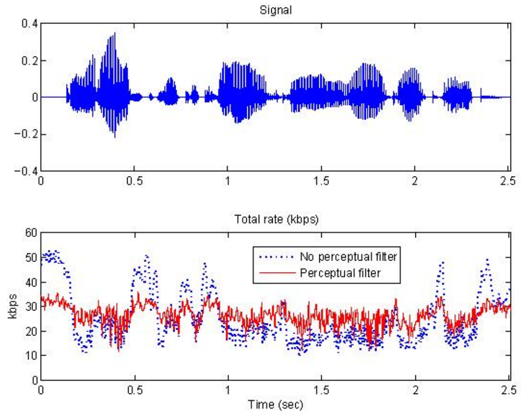

Fig. 3. Rate variation (for CE quantization).

Thus, we see that in both the CR and CE cases, and under high-rate assumptions, the rate depends linearly on the log of the distortion with the scale factor $K/2$ and with an additive constant which depends only on data and on modeling.

We compared the practical performance of the proposed coding scheme with the theoretical performance predicted under high-rate assumptions (Eqs. (17) and (18)) in both CR and CE cases. We used the test material described in section 4.1. The AR model parameters (i.e., LPC and variance) were not quantized. Signal quantization was performed at different rates (from 1 to 8 bits per sample). The distortion was measured as MSE in the gain-normalized (Sec. 3.2) domain.

The results are plotted on figure 2 together with the theoretical rate-distortion curves given by equations (17) and (18). We see that, starting from the rate of 3 bits per sample, the practical coder performance for both CR and CE cases follows the theoretical performance predicted under high-rate assumptions.

4.3. Rate variation

Figure 3 presents the rate variation of the CE coder for one wideband speech sentence in two configurations: (i) without the perceptual filter (i.e., as described in the paper), and (ii) using the AMR-WB [1] perceptual filter (see, e.g., [8]). The average rate is about 27 kbps. One can note that the rate variation is reasonable, and has a smaller variance when the perceptual filter is applied, which should be done in a practical coder.

5. CONCLUSION

We conclude that it is possible to build a practical flexible coding scheme that is able to adapt in real time to any particular rate, and has low computational complexity and low storage requirements independently of the rate, with a performance equivalent to training-based fixed-rate CELP at any rate. The performance of the proposed high-rate theory-based scheme follows the theoretically predicted performance starting form the rate of three bits per sample, and is close to the high-rate-theory derived rate-distortion relation at lower rates. The rate variation of the variable-rate coder (based on the CE quantization) is reasonable, and does not require significant flexibility of the transmission channel.

6. REFERENCES

[1] 3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec," 2005, technical specification.

[2] R. M. Gray, *Source coding theory*. Kluwer Academic Press, 1990.

[3] A. Subramaniam and B. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 2, pp. 130–142, March 2003.

[4] D. Zhao, J. Samuelsson, and M. Nilsson, "GMM-based entropy-constrained vector quantization," in *IEEE ICASSP'07*, vol. 4, 15-20 April 2007, pp. IV–1097–IV–1100.

[5] W. B. Kleijn and K. K. Paliwal, Eds., *Speech Coding and Synthesis*. Amsterdam, The Netherlands: Elsevier, 1995.

[6] M. Y. Kim and W. B. Kleijn, "KLT-based adaptive classified VQ of the speech signal," *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 3, pp. 277–289, May 2004.

[7] J. Samuelsson, "Waveform quantization of speech using Gaussian mixture models," in *IEEE ICASSP '04*, vol. 1, May 2004, pp. 165–168.

[8] M. Li and W. B. Kleijn, "A low-delay audio coder with constrained-entropy quantization," in *IEEE WASPAA'07*, Oct. 2007.

[9] W. B. Kleijn and A. Ozerov, "Rate distribution between model and signal," in *IEEE WASPAA*, Mohonk, NY, Oct. 2007.

[10] J. Samuelsson, "Toward optimal mixture model based vector quantization," in *ICICS'05*, Dec. 2005, pp. 1329–1333.

[11] J.-H. Chen, "High-quality 16 kb/s speech coding with a one-way delay less than 2 ms," in *IEEE ICASSP'90*, vol. 1, April 1990, pp. 453–456.

[12] W. B. Kleijn, "A basis for source coding," Nov. 2006, lecture notes KTH, Stockholm.