Spectral Quantization for Wideband Speech Coding

Giuseppina Biundo, Sara Grassi, Michael Ansorge, Fausto Pellandini

***Abstract*---- In this paper we describe the design of a spectral quantization scheme for wideband speech coding. The theoretical basis is given, together with experimental design, and implementation in a proprietary CELP (Code Excited Linear Prediction) coder.**

***Index Terms*-- Wideband speech coding, spectral quantization.**

I. Introduction

Source coding of wideband (0.05-7.0 kHz) speech signals is currently a predominant research topic [1], driven by standardization activities at ETSI (European Telecommunications Standards Institute) and ITU-T (International Telecommunications Union – Telecommunications Standards Sector). In comparison to narrowband speech coding, for which the speech signal is band-limited to 0.3-3.4 kHz and sampled at 8kHz, wideband speech coding improves naturalness and intelligibility of the decoded speech, at the cost of a doubled sampling frequency (f_s = 16 kHz). Wideband speech coding finds applications in video conferencing, wireline and mobile telephony, circuit switched and packet networks, and multimedia broadcast. Besides, wideband speech coders could improve human-machine interfaces and recognition over coded speech.

Spectral quantization in narrowband speech coding has been extensively studied and there exists a lot of scientific literature on the subject. Examples are found in a large amount of standards such as the ITU-T G.723.1 and G.729, and GSM (Global Systems for Mobile communications) EFR (Enhanced Full-Rate), HR (Half-Rate) [2] and NB-AMR (Narrow Band Adaptive Multi-Rate) [3]. On the other hand, the topic of spectral quantization for wideband speech coders is newer, with lesser amount of information available.

This paper is organized as follows: Section II reviews some basis in the design of spectral quantization for wideband speech coders. Section III discusses some examples found in literature and standards. Experimental design of different spectral quantization schemes is given in Section IV. The finally chosen quantization scheme is included in a CELP coder as explained in Section V. Conclusions and further work are given in Section VI.

The authors are with the Institute of Microtechnology, University of Neuchâtel, Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland. Phone: +41 32 718 34 32; Fax: +41 32 718 34 02; e-mail: Giuseppina.Biundo@unine.ch.

II. Spectral quantization

A. Spectral quantization parameters [5][23]

In Linear Prediction (LP) - based speech coders, the spectral envelope of a speech frame is modeled with the all-pole filter $1/A_m(z)$ given by:

$$A_m(z) = \sum_{i=1}^{m} a_i \cdot z^{-i} \qquad (1)$$

where the a_i are the LP coefficients and m is the order of the model, which is typically 10 in narrowband coders, and 16 in wideband coders. An order of 16 is used throughout the whole paper.

In forward LP-based coders, the spectral information, contained in the a_i, is quantized and transmitted. As the a_i are difficult to quantize, different one-to-one representations such as Parcor Coefficients (k_i), Log-area-ratio, Line Spectrum Pairs (LSP) [5] and Immittance Spectral Pairs (ISP) [6] are used. We focus on LSP and ISP, as LSPs are widely used in recent narrowband standards as well as most speech coders in the scientific literature, whereas ISPs are used in the new ETSI WB-AMR (Wide Band Adaptive Multi-Rate) speech coder [4].

1) LSP and ISP

LSP, also referred to as Line Spectrum Frequency (LSF) is a spectral envelope representation with good quantization properties, such as bounded range, intra- and inter-frame correlation, localized spectral sensitivity and simple check of filter stability. The LSPs are calculated from $A_m(z)$ as follows: the symmetric polynomials $P(z)$ and $Q(z)$ are given by:

$$\begin{aligned} P'_m(z) &= A_m(z) + z^{-(m+1)} A_m(z^{-1}) = (1+z^{-1}) \cdot P(z) \\ Q'_m(z) &= A_m(z) - z^{-(m+1)} A_m(z^{-1}) = (1-z^{-1}) \cdot Q(z). \end{aligned} \qquad (2)$$

$P(z)$ and $Q(z)$ are completely specified by the angular position of their roots in the upper semicircle of the z-plane. These angles are the m LSP parameters, denoted by ω_i. Hereafter we use the LSPs in the frequency domain, with $f_i = f_s \cdot \omega_i / 2\pi$. If $1/A_m(z)$ is stable, the LSPs are ordered as in Eq. (3). The converse is also true, namely if an LSP set is ordered, its corresponding filter $1/A_m(z)$ is stable [5].

$$0 < f_1 < f_2 < \ldots < f_m < f_s/2 \qquad (3)$$

Immittance spectral pairs (ISP) have slightly better quantization properties and lesser computational complexity. They are used in the new ETSI WB-AMR speech coder. Similarly to LSP, the symmetric polynomials $P(z)$ and $Q(z)$ are given by:

$$\begin{aligned} P'(z) &= A_m(z) + z^{-m} A_m(z^{-1}) = (1 + a_m) \cdot P(z) \\ Q'(z) &= A_m(z) - z^{-m} A_m(z^{-1}) = (1 - a_m) \cdot (1 - z^{-2}) \cdot Q(z) \end{aligned} \tag{4}$$

the angular positions of their roots (upper semicircle of the z-plane) are the first $m-1$ ISPs. Note that the first $m-1$ ISPs of a system of order m are the LSPs of the system of order $m-1$. The mth ISP is derived from the last LP coefficient a_m (which is equal to the last Parcor coefficient $k_m : -1 < k_m < 1$) [7].

B. Performance measures [8]

The Spectral Distortion (SD), expressed in dB, of a speech frame is:

$$SD = \sqrt{\frac{100}{(f_2 - f_1)} \int_{f_1}^{f_2} \left[\log_{10} S(f) - \log_{10} Sq(f)\right]^2 df} \tag{5}$$

where f is the frequency in Hz, f_1 and f_2 specifying the frequency range, $S(f)$ and $Sq(f)$ being the original and quantized spectrum of $1/A_m(z)$. In case of narrowband coders, a widely accepted criterion is the following [8]: the quantization process is considered “transparent” (i.e. does not introduce audible distortion) if, over a large number of frames, the average SD is less than 1 dB, and there are less than 2% of outliers with SD lying in the range of 2-4 dB, and there are no outliers with SD larger than 4 dB. As it is done in [9], we decided to use the same criterion for wideband coders, with a frequency range of 0-7 kHz, coupled with a listening test verification of the system depicted in Fig. 1. A first (not yet widely accepted) attempt to establish the requirements for transparency in wideband speech coders is given in [10].

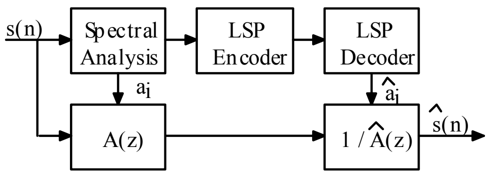

Fig. 1. System used for subjective listening tests.

C. Quantization techniques [11], [9]

Vector Quantization (VQ) effectively exploits the correlation between neighboring LSPs (intra-frame correlation) for bit-rate reduction. Sub-optimal VQ, such as Split Vector Quantization (SVQ) [8] and Multi-Stage Vector Quantization (MSVQ) [12], is used to decrease the storage and complexity of the full-search VQ. In SVQ the LSP vector is partitioned into smaller sub-vectors, and each sub-vector is quantized using a full-search VQ. In practice, SVQ is preferred for LSP quantization, as the localized spectral sensitivity of the LSPs limits the spectral distortion leakage from one region to the other. There exist also quantization schemes which combine SVQ and MSVQ. The experimental design presented in the next section deals only with SVQ. Besides VQ techniques which exploit the intra-frame correlation of the LSPs, predictive VQ (PVQ) techniques exploit the temporal correlation of consecutive LSP vectors (inter-frame correlation) [9].

III. EXAMPLES OF SPECTRAL QUANTIZATION

A short revision of previous work on spectral quantization for wideband speech coders is given below.

- In [9], Guibe & al. give an overview of different wideband speech coders and quantization techniques. Their aim is to provide a comparative study of various quantizers like predictive, memoryless and safety-net (combination of predictive and memoryless) vector quantizers as well as split, multi-stage quantizers, or split-multi-stage vector quantizers (S-MSVQ) which combine the two last techniques.
- In [13], Gibbs & al. consider the impact of tandemming wideband and narrowband speech codecs, on performances of split (18-order) LSP VQ with Moving Average (MA) prediction of order 0, 1 and 2. The optimal bit allocation in the 6 split codebooks, embedding each 3 LSPs, is described for wideband speech.
- In [10], Ferhaoui & al. discuss multi-stage VQ and propose a new spectral distortion measure criterion for transparency in wideband speech coding: the quantization process is termed “transparent” if the average SD is less than 1.6 dB, if there are less than 4% of outliers with SD within 3-5 dB, and if there are no outliers with SD larger than 5 dB.
- Finally, the only example of wideband LPC-based standard (known to the authors) is described in the next sub-section.

A. Spectral quantization in the ETSI WB-AMR [4]

The new ETSI WB-AMR uses a sub-band coding approach. A low pass filtered speech signal (0.0-6.4 kHz) is obtained by decimation. A 16 order LPC extraction is realized on 20 ms frames of filtered speech. The computed LP coefficients are converted to ISPs, as explained in Section II, and a residual ISP vector is extracted by 1st order MA prediction. Finally, the residual ISP vector is quantized using an S-MSVQ.

WB-AMR uses 9 modes, each with a different bit-rate ranging from 6.6 to 23.85 kbps, for adaptation to different channel conditions. The ISPs quantization scheme requires a bit-rate contribution of 46 bits per 20 ms frame in all modes with the exception of the lowest bit-rate, which requires only 36 bits.

The residual ISP vector is first split into two sub-vectors $\mathbf{r}_1(n)$ and $\mathbf{r}_2(n)$ of dimensions 9 and 7. These two sub-vectors are quantized in two stages. In the first stage, $\mathbf{r}_1(n)$ and $\mathbf{r}_2(n)$ are quantized with 8 bits each. The quantization in the second stages differs according to the mode. For all modes with the exception of the lowest bit-rate, the resulting error vectors are split into 3 and 2 sub-vectors, of dimension 3-3-3 and 3-4, respectively. They are quantized using 6, 7, 7 and 5, 5 bits. For the lowest bit-rate, only the first resulting error vector is split into 2 vectors of dimension 5-4, quantized using 7, 7 bits. The second resulting error vector is not split and has thus a dimension of 7. It is quantized using 6 bits.

IV. EXPERIMENTAL DESIGN

A. *LSP databases for training and testing*

The training and testing data are built up using three different speech databases: the TIMIT [14] (US English), the BDSONS [15] (French), and the ITU multi-Lingual Speech Database for Telephonometry (rec. P.50 annex I) [16]. The LSP database for training is generated from the following speech data: the first one hour of the TIMIT training sub-set, the whole ITU (59'), and one hour of speech randomly chosen from the BDSONS CD-ROM 6/7. The LSP database for testing is generated from the BDSONS CD-ROM 3/7, sub-directory LABIS (a short story in French uttered by 30 speakers).

The LSPs are first calculated from the speech data, as it is done in the proprietary Wideband CELP coder (cf. Section V), and the LSP vectors corresponding to silence frames are then removed (in the training database). This was preferred to doing silence removal before LSP calculation, in order to avoid having non-contiguous speech within the LSP calculation frames (situation that we consider artificial).

B. *SVQ split and optimal bit allocation*

The first issue in split VQ design is defining the best partition of the 16-component LSP vector. Complexity decreases with the number of splits, concomitant to the VQ performance. In [9] it is argued that the 3 higher frequency LSPs have a different statistical behavior than the rest, leading to a 13-3 split, for a 16-components LSP vector. Then, they further split the lower frequency LSP vector, obtaining a 6-7-3 split. In [13] it is argued that a common choice in narrowband coding is to use 3 components per split giving a 3-3-3-3-3-3 split for an 18-component LSP. Accordingly, we decided to use five codebooks with dimension 4-3-3-3-3. All experiences presented in this paper use this split, but further work is being carried out on this topic.

In order to determine the optimal bit allocation for each codebook, we start with a 15-bit budget, 3 bits per codebook. At every stage of the optimization, the bit budget is increased by one bit, and this extra bit is tentatively allocated to each codebook, training and testing is performed, spectral distortion is measured, and the configuration that gives the best marginal improvement is chosen. This is done for a total number of bits ranging from 16 to 45. The optimal bit allocation is given in TABLE I, along with the average spectral distortion and percentage of outliers. In order to reduce the research complexity, only one set of quantized LSPs out of five is used in the testing. According to the results, we focus on a bit budget of 34 and 45 bits for the continuation of the experiments. 34-bit allocation corresponds to the wideband transparency proposed by Ferhaoui & al. [10], and 45 bits to the commonly accepted narrowband transparency criterion. The codebooks were trained using LBG (Linde-Buzo-Gray) algorithm [17]. Euclidean distance was used for training and testing. Verification listening tests on the system depicted in Fig. 1 with the testing database show that the 45-bit configuration is transparent while the 34-bit configuration involves significant distortion. The 45-bit configuration corresponds to a quantization complexity of 896k multiplications and 1495k additions per second, while requiring a memory space of 43 KBytes, each stored value occupying 2 Bytes.

TABLE I

OPTIMAL BIT RATE ALLOCATION FOR EACH OF THE FIVE CODEBOOKS WITH SUB-VECTOR DIMENSIONS 4-3-3-3-3, ALONG WITH THE AVERAGE SPECTRAL DISTORTION AND PERCENTAGE OF OUTLIERS FOR A BIT BUDGET RANGING FROM 16 TO 45.

Bit budget	*Bit allocation in codebooks 1 to 5*	*Average Spectral Distortion (dB)*	*Frames with 2<SD≤4 dB in %*	*Frames with SD>4 dB in %*
16	(4,3,3,3,3)	3.2116	85.89	11.85
17	(5,3,3,3,3)	3.0491	86.45	9.28
18	(5,4,3,3,3)	2.9166	87.76	6.16
19	(5,4,4,3,3)	2.8064	89.16	3.65
20	(6,4,4,3,3)	2.7100	87.14	2.84
21	(7,4,4,3,3)	2.5901	82.42	2.12
22	(7,4,5,3,3)	2.4929	78.21	1.51
23	(7,5,5,3,3)	2.3907	74.56	0.94
24	(7,6,5,3,3)	2.2827	67.37	0.58
25	(7,6,5,4,3)	2.1904	62.10	0.18
26	(7,6,5,5,3)	2.1129	56.19	0.18
27	(7,6,6,5,3)	2.0360	49.8	0.09
28	(7,7,6,5,3)	1.9543	42.39	0.03
29	(8,7,6,5,3)	1.8741	34.69	0.05
30	(8,7,7,5,3)	1.8104	29.58	0.04
31	(8,7,7,5,4)	1.7439	23.38	0.01
32	(8,7,7,6,4)	1.6807	18.53	0.02
33	(9,7,7,6,4)	1.6046	14.34	0.01
34	(9,8,7,6,4)	1.5430	10.78	0.00
35	(10,8,7,6,4)	1.4815	8.40	0.02
36	(10,9,7,6,4)	1.4268	6.27	0.00
37	(10,9,8,6,4)	1.3700	4.83	0.00
38	(10,9,8,7,4)	1.3178	3.73	0.00
39	(11,9,8,7,4)	1.2638	2.83	0.00
40	(11,9,8,8,4)	1.2186	2.50	0.00
41	(11,9,8,8,5)	1.1748	1.32	0.00
42	(11,9,9,8,5)	1.1251	0.91	0.00
43	(11,9,9,8,6)	1.0801	0.73	0.00
44	(11,10,9,8,6)	1.0354	0.50	0.00
45	(12,10,9,8,6)	0.9889	0.35	0.00

C. *Prediction order and coefficients [18]*

The quantization used in the previous sub-section is referred to as *memoryless* SVQ as no temporal information from previous frames is used in the quantization of the current frame. Although memoryless VQ improves robustness against channel errors, *predictive* SVQ exploits the temporal correlation of consecutive LSP vectors (inter-frame correlation).

The prediction can be done either by auto regressive (AR) or by moving average (MA) filter. Usually, MA prediction requires a higher order predictor to reach the same performance as with AR prediction. The main advantage of the MA system is the finite impulse response of the prediction filter which leads to limited bit error propagation in the case of noisy channel transmission.

In the MA case, the ith linearly predicted value, $\hat{f}_i(n)$, is given by [18]:

$$\hat{f}_i(n) = \sum_{k=1}^{q} b_i(k) \cdot \tilde{e}_i(n-k) \tag{6}$$

where the $e_i(n-k)$ are the previously coded prediction errors, $b_i(k)$ are the prediction coefficients, and M is the predictor order.

The MA prediction coefficients can be determined either by an LMS (Least-Mean-Square) algorithm [19] or by a standard method based on a high order AR approximation of the MA process [20]. The decomposition theorem due to Wold (1938) asserts that any MA process can be represented uniquely by an AR model of possibly infinite order [21]. If we let the MA process of order q, denoted by MA(q), be modeled by an AR model of order p, denoted by AR(p), where $p \gg q$, then the q order MA filter $B(z)$ can be expressed as a function of the AR filter $A(z)$ following $B(z) = 1/A(z)$, with

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \quad B(z) = \sum_{k=0}^{q} b_k z^{-k}. \tag{7}$$

Thus the parameter sets $\{b_k\}$ and $\{a_k\}$ are inter-related by:

$$\hat{a}_n + \sum_{k=1}^{q} b_k \hat{a}_{n-k} = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \tag{8}$$

where $\{\hat{a}_k\}$ are the parameters obtained by fitting the data to an $AR(p)$ model. A better fit is obtained using a least-squares error criterion and minimizing the squared error ε specified in Eq. (9) for selecting the parameters $\{b_k\}$:

$$\begin{aligned} \varepsilon &= \sum_{n=0}^{p} \left[\hat{a}_n + \sum_{k=1}^{q} b_k \hat{a}_{n-k} \right]^2 - 1; \\ \hat{a}_0 &= 1, \quad \hat{a}_k = 0 \quad if \quad k < 0. \end{aligned} \tag{9}$$

This leads to a very simple form, namely the resolution of a q^{th}-order Levinson Durbin equation using the parameters $\{\hat{a}_k\}$ as input data.

We decided to use the MA prediction due to its robustness to channel transmission errors. We tested the MA prediction of order 0 (no prediction), 1, 2 and 3. Although slightly sub-optimal, the MA predictors were trained in open-loop for the purpose of this work, and then maintained for all quantization tests, in order to reduce the overall complexity of the investigations. Moreover, we decided to use the mean value of the MA coefficients extracted within each split, as we observed that these values are very close.

The spectral distortion was measured, for the 34- and 45-bit schemes, using prediction order of 0, 1, 2 and 3. Results are plotted in Fig. 2 (cf. curves a) and c)). We can observe that there is a considerable gain in using MA prediction of order one and little gain in increasing the prediction order beyond one. Similar results were obtained when perceptual weighting is included in the quantization, as explained in the next sub-section. For simplicity, after determining the MA prediction coefficients, the VQ was trained by an open-loop procedure.

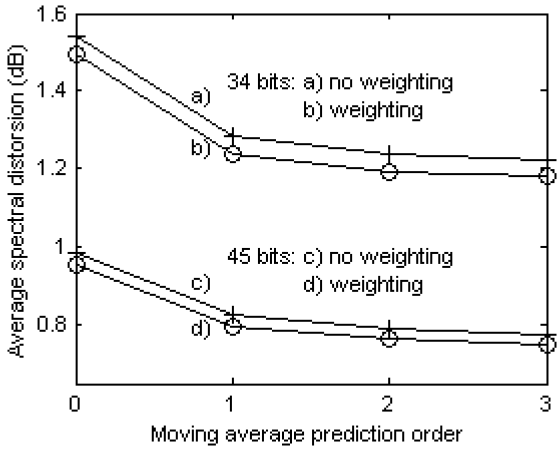

Fig. 2. Spectral distortion for different MA orders, 34- and 45-bit schemes, with and without weighting.

D. Perceptual weighting

The Eucledian distance measure used in the previous experiments for training and quantization assigns an equal weight to each component of the LSP vector $\mathbf{f}$, and is thus not taking into account the human auditory perception properties. The weighted Euclidean distance, $d(\mathbf{f},\hat{\mathbf{f}})$, measured between $\mathbf{f}$ and its quantized value $\hat{\mathbf{f}}$, assigns the weight w_i to each LSP f_i, according to its perceptual importance :

$$d(\mathbf{f},\hat{\mathbf{f}}) = \sum_{i=1}^{m} w_i \left(f_i - \hat{f}_i \right)^2. \tag{10}$$

The LSPs in the region of spectral peaks (formants) should be weighted more than those on the spectral valleys, and more weight should be assigned to LSPs corresponding to high-amplitude formants than to those corresponding to low-amplitude formants. The LSPs display a cluster pattern around the formants. A cluster of (2 to 3) LSPs characterizes a formant frequency and the bandwidth of the given formant depends on the closeness of these LSPs. We use a modification of the ITU-T Rec. G.729 weighting scheme specified by :

$$w_i = \begin{cases} \alpha & \text{if } d_i > \alpha, \\ 10 \cdot (d_i - \alpha)^2 + \alpha & \text{otherwise;} \end{cases}$$

$$\alpha = 1.0;$$

$$d_i = \omega_{i+1} - \omega_{i-1}; \quad \omega_0 = 0.04 \cdot \pi; \quad \omega_{11} = 0.92 \cdot \pi.$$

The modification takes into account the greater density of LSPs in the case of wideband coding: We use, for the weighting calculation, the following parameterized function of the LSP distance d_i :

$$w_i = \begin{cases} \alpha & \text{for } d_i > (1-\alpha)K, \\ \dfrac{1}{(1+\alpha)}\left[\dfrac{1}{K^2}(d_i - K)^2 + \alpha\right] & \text{otherwise;} \end{cases}$$

$\alpha = 0.1, \quad \text{K=0.6;}$

$d_i = \omega_{i+1} - \omega_{i-1}; \quad \omega_0 = 0; \quad \omega_{17} = \pi.$

The parameters α and K were experimentally tuned, giving $\alpha = 0.1$ and $K = 0.6$. Note that with α=0.1 and K=0.9, our parameterized function is almost the same as the one of G.729. Spectral distortion is measured for the 34- and 45-bit schemes, using prediction order of 0, 1, 2 and 3. The codebooks are trained using LBG algorithm and Euclidean distance. The quantization (testing) is done using the weighted Euclidean distance. Results are plotted in Figure 2 (cf. curves b) and d)). The use of perceptual weighting improves the spectral distortion by approximately 0.04 and 0.03 dB for the 34, and the 45-bit schemes respectively.

In order to observe the effect of the weighting, we plot the SD, in the form of a histogram, for the 1st order MA prediction vector quantizer at 34 bits/frame, using un-weighted and weighted Euclidean distance measure for quantization. These plots are shown in Fig. 3.

Experiments showed no further significant improvement when using the weighted measure also during the training phase.

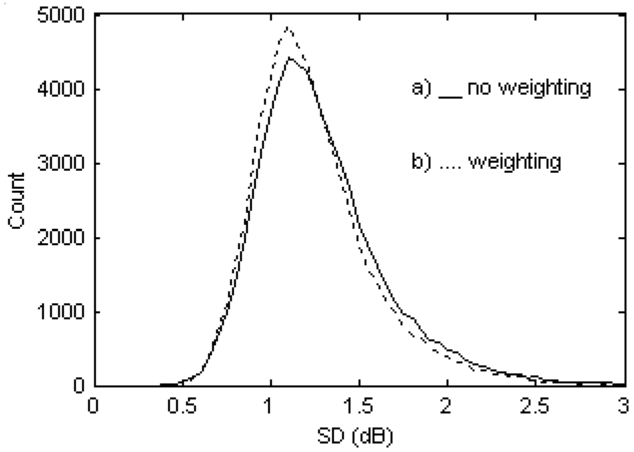

Fig. 3. Spectral distortion (SD) histogram for the 34 bits/frame 1st order prediction split-vector quantizer using the un-weighted and weighted Euclidean distance measure for quantization.

E. Second bit allocation iteration

In order to determine the optimal bit allocation in the case of 1st order prediction and quantization using the weighted Euclidean distance measure, we repeat the procedure for bit allocation explained in subsection B. The obtained optimal bit allocation is given in TABLE II. According to the results, we focus on a bit budget of 40 bits, which corresponds to the commonly accepted narrowband transparency criterion. Notice that the transparency criterion is now reached with 5 bits less.

F. The final choice

The finally chosen quantization scheme uses predictive split VQ with a 4-3-3-3-3 split and 1st order MA prediction. Training of the codebooks is done with the LBG algorithm and Euclidean distance. Quantization is done with weighted Euclidean distance. The bit budget is 40, with the bit allocation configuration of TABLE II. This 40-bit configuration corresponds to a quantization complexity of 1098k multiplications and 949k additions per second, and to 21.5 KBytes of memory space, each stored value still occupying 2 Bytes. We have tested if the scheme could produce unstable (i.e. not ordered) LSPs. We observed that less than 2 % of LSP vectors had to be rearranged and that about 5 % of successive LSPs were too close and needed to be shifted. A test was thus added to reorder the LSPs, and to separate by 50 Hz those which were too close. Applying verification listening tests on the system depicted in Fig. 1 finally showed that the chosen scheme is transparent for the used testing database. The output quality is generally the same as the one obtained with 45 bits, without MA prediction and without weighting the distortion measure during quantization (cf. subsection IV.B). Nevertheless, in some very rare cases, the output quality is slightly degraded.

V. Wideband CELP Speech Coder

The designed spectral quantization scheme was included in a proprietary variable rate ACELP (Algebraic CELP) wideband coder. This coder uses a full-band approach, 16 kHz sampling rate and a 20 ms frame size, with four 5 ms sub-frames. The bit-rate distribution is given in TABLE III, for the case of the finally chosen 40-bit spectral quantization scheme. A 16th order LP analysis is performed once per frame. The autocorrelation coefficients processed over 30 ms speech windows are converted to LP coefficients using the Levinson-Durbin algorithm. The LP coefficients are then in turn transformed to the LSP domain for quantization and interpolation purposes. A further detailed description of the functional blocks of the coder is not possible in the frame of this paper.

VI. Conclusions and Future Work

We have investigated the design of spectral quantization for a wideband multi-rate CELP coder. Different 16- to 45-bit memoryless SVQ schemes were first designed and tested. Based on the achieved results, the experiments were then focused on the 34- and 45-bit budget schemes, investigating the MA prediction of order 0 to 3, and also the weighted Euclidean distance measure for quantization. The optimal bit allocation was then determined for the case of 1st order prediction and quantization using the weighted Euclidean distance measure. This lead to the final 40-bit scheme fulfilling the “transparency” criterion, with almost no audible distortion when applying listening verification tests using the system depicted in Fig. 1.

Future work will be in optimizing the number of splits as well as the split size for the SVQ. The multi-stage quantization scheme, and the combined split and multistage VQ will be considered for this purpose, as well as a classification scheme featuring a prediction limited to the voiced speech regions only.

TABLE II

OPTIMAL BIT RATE ALLOCATION FOR EACH OF THE FIVE CODEBOOKS WITH DIMENSION 4-3-3-3-3, ALONG WITH THE AVERAGE SPECTRAL DISTORTION AND PERCENTAGE OF OUTLIERS FOR A BIT BUDGET RANGING FROM 16 TO 45 AND FOR 1ST ORDER PREDICTION AND QUANTIZATION USING THE WEIGHTED EUCLIDEAN DISTANCE MEASURE.

Bit budget	*Bit allocation in codebooks 1 to 5*	*Average Spectral Distortion (dB)*	*Frames with 2<SD≤4 dB in %*	*Frames with SD>4 dB in %*
16	(4,3,3,3,3)	2.7904	85.66	4.65
17	(4,4,3,3,3)	2.6437	82.93	3.01
18	(5,4,3,3,3)	2.5097	78.01	2.02
19	(5,4,4,3,3)	2.3749	72.06	1.26
20	(6,4,4,3,3)	2.2696	64.33	0.90
21	(6,5,4,3,3)	2.1629	57.25	0.57
22	(7,5,4,3,3)	2.0739	49.05	0.45
23	(7,5,4,4,3)	1.9832	41.93	0.24
24	(7,5,5,4,3)	1.8936	34.78	0.13
25	(7,6,5,4,3)	1.8083	28.42	0.10
26	(8,6,5,4,3)	1.7342	23.26	0.05
27	(8,6,6,4,3)	1.6623	19.35	0.04
28	(8,6,6,5,3)	1.5905	14.55	0.03
29	(8,7,6,5,3)	1.5241	11.37	0.02
30	(9,7,6,5,3)	1.4643	9.43	0.02
31	(9,7,6,6,3)	1.4084	7.60	0.00
32	(9,7,7,6,3)	1.3515	6.06	0.00
33	(9,7,7,6,4)	1.2956	4.59	0.00
34	(9,8,7,6,4)	1.2431	3.60	0.00
35	(10,8,7,6,4)	1.1956	2.69	0.00
36	(10,8,7,7,4)	1.1484	2.19	0.00
37	(10,8,8,7,4)	1.1043	1.75	0.00
38	(10,9,8,7,4)	1.0626	1.46	0.00
39	(11,9,8,7,4)	1.0224	1.14	0.00
40	(11,9,8,7,5)	0.9815	0.73	0.00
41	(11,9,8,8,5)	0.9451	0.63	0.00
42	(11,9,9,8,5)	0.9075	0.46	0.00
43	(11,10,9,8,5)	0.8588	0.25	0.00
44	(11,10,9,8,6)	0.8353	0.28	0.00
45	(12,10,9,8,6)	0.7976	0.21	0.00

TABLE III

BIT ALLOCATION FOR VARIABLE RATE ACELP WIDEBAND CODER FOR A 20 MS FRAME.

Mode	Parameter	1st Sub-frame	2nd Sub-frame	3rd Sub-frame	4th Sub-frame	Total per frame
All	LSP set					40
	Pitch lag	9	6	9	6	30
	Gains	7	7	7	7	28
	Sub-total					**98**
13.9 kbits/s	ACELP pulses	45	45	45	45	180
	Total					**278**
17.9 kbits/s	ACELP pulses	65	65	65	65	260
	Total					**358**
20.9 kbits/s	ACELP pulses	80	80	80	80	320
	Total					**418**

VII. ACKNOWLEDGEMENTS

This work was partially supported by the Swiss Federal Office for Education and Science under Grant OFES C00.0105 (COST 276 project).

Special thanks are addressed to Dr. Benito Carnero for his contribution to sub-section IV.D.

VIII. REFERENCES

[1] J. Schnitzler and P. Vary, "Trends and perspectives in wideband speech coding", *Signal Processing*, Elsevier, The Netherlands, Vol. 80, no. 11, pp. 2267-2281, Nov. 2000.

[2] L. Hanzo, F. C. A. Somerville and J. P. Woodard, "Standard forward-adaptive CELP codecs", Chapter 7, in [22].

[3] 3G TS 26.090 V3.1.0 (1999–12) document, in ftp://ftp.3gpp.org/Specs/2000-09/R1999/26_series/ (November 14, 2001).

[4] 3GPP TS 26.190 V5.0.0 (2001–03) document, in ftp://ftp.3gpp.org/Specs/2001-09/Rel-5/26_series/ (November 14, 2001).

[5] L. Hanzo, F. C. A. Somerville and J. P. Woodard, "Speech spectral quantization", Chapter 4, in [22].

[6] Y. Bistritz and S. Peller, "Immittance spectral pairs (ISP) for speech encoding", *Proc. 1993 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-93)*, Minneapolis, Vol. 2, pp. 9-12, April 1993.

[7] C.-F. Chan, "Efficient quantization of LPC parameters using a mixed LSP/PARCOR representation", *Signal Processing VII: Theories and Applications*, M. Holt & Co. Eds., pp. 939-942, 1994.

[8] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters", Chapter 12, in [23].

[9] G. Guibe, H. T. How and L. Hanzo, "Comparative study of wideband speech spectral quantization schemes", *Proc. 3rd ITG Conference Source and Channel Coding*, pp. 181-186, Munich, Germany, Jan. 2000. Also in [22] (Section 4.4).

[10] M. Ferhaoui and S. Van Gerven, "LSP quantization in wideband speech coders", *Proc. of 1999 IEEE Workshop on Speech Coding*, pp. 25-27, Porvoo, Finland, June 1999.

[11] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Vector quantization for speech transmission", Chapter 9, in [23].

[12] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Theory for transmission of vector quantization data", Chapter 10, in [23].

[13] J.A. Gibbs and J.M. Hoskin, "LSP split vector quantization for wideband codecs with narrowband tandemming", *Proc. of 2000 IEEE Workshop on Speech Coding*, Delavan, Wisconsin, USA, pp. 120-122, Sept. 2000.

[14] J. Garofol and al., *Darpa TIMIT, Acoustic-Phonetic Continuous Speech Corpus CD-ROM*, National Institute of Standards and Technology, NISTIR 493, USA, Oct. 1990.

[15] *Bdsons, Base de Données des Sons du Français*, CD-ROM, (CD no. 1) edited by Jean-François Serignat and Ofelia Cervantes, ICP, Grenoble (France) ; published by CEDROM Technologies, 30 avenue de l'Observatoire, 75015 Paris, France.

[16] http://www.itu.int/publications/itu-t/list-t-soft.html (November 14, 2001).

[17] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design", *IEEE Transactions on Communications*, vol. 28, pp. 84-95, Jan. 1980.

[18] J. Skoglund and J. Linden, "Predictive VQ for noisy channel spectrum coding: AR or MA?", *Proc. 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-97)*, Munich, Germany, Vol. 2, pp. 1351 - 1354, April 1997.

[19] H. Ohmuro, T. Moriya, K. Mano, and S. Miki, "Vector quantization of LSP parameters using moving average interframe prediction", *Electronics and Communications in Japan,* Part3, Vol. 77, No. 3, pp. 12-26, March 1994.

[20] S. L. Marple, Jr., *Digital spectral analysis with applications*, Prentice-Hall International, Inc., 1987.

[21] J. G. Proakis and D. G. Manolakis, "Power spectrum estimation", Chapter 12, *Digital signal processing, principles, algorithms, and application*, Third Edition, Prentice-Hall International, Inc., 1996.

[22] L. Hanzo, F. C. A. Somerville and J. P. Woodard, *Voice Compression and Communications*, IEEE Series on Digital & Mobile Communication, John Wiley & Sons, Inc., Publication, NY, USA, 2001.

[23] W. B. Kleijn and K. K. Paliwal (Editors), *Speech Coding and Synthesis*, Elsevier, Amsterdam, The Netherlands, 1995.