FLEXIBLE SPECTRUM CODING IN THE 3GPP EVS CODEC

Adriana Vasilache[*]*, Anssi Rämö*[*]*, Hosang Sung*[**]*, Sangwon Kang*[***]*, Jonghyeon Kim*[***]*, and Eunmi Oh*[**]

[*] Nokia Technologies, Nokia, Finland
[**] DMC R&D Center, Samsung Electronics Co. Ltd., Korea
[***]Dept. of Electronics and Communication Eng., Hanyang University, Korea

ABSTRACT

This paper proposes a flexible encoding technique based on multistage multiple scale lattice vector quantization and block-constrained trellis coded vector quantization. It is used for the spectrum encoding, more precisely encoding of the LSF parameters, and incorporated in the recently standardized 3GPP EVS codec. The proposed method can handle multiple bit allocations and signal types with low complexity and low memory requirements.

Index Terms— Speech coding, LSF quantization, lattice quantization, block constrained trellis coding

1. INTRODUCTION

The CELP (code excited linear prediction) paradigm is still one of the most efficient techniques used for coding of speech signals. It combines the use of the parameters of linear predictive model of the speech signal with analysis by synthesis methods. The reconstructed and original speech are compared, and the excitation parameters are adjusted to minimize the difference, before the code is transmitted. The encoding of the linear prediction coefficients (LPC) is not done in LPC representation because even small quantization errors in the coefficients can lead to significant differences in the quantized spectrum whose stability is not guaranteed. An alternative representation like the line spectrum frequency (LSF) is a better choice because the filter stability can be ensured by simply checking that the LSF parameters are in increasing order [1]. Another advantage of the LSF representation is that parameters sensitivity is relatively uniform across the spectrum [2].

The necessary bitrates for transparent LPC quantization should be above 2 bits per sample, depending on the signal or whether prediction is used or not and how strong it is. During the last 20-30 years, there has been extensive work on spectrum quantization and ways to provide alternatives to unstructured optimized vector quantizers that, at the considered bitrate, would have too many codevectors to be of practical use. Split vector quantizers [3], [4], multistage vector quantizers [5], [6], [7] have been used to alleviate the storage and complexity burden of the optimized VQ. Lattice quantizers have also been proposed in conjunction with split VQ [8], with multistage VQ [9], [10], or as standalone quantizers [11] for further reductions of complexity and storage requirements. Arguing that training optimized vector quantizer codebooks, especially for high number of codevectors, is sensitive to the training data, several solutions based on mixture of various distributions have been recently proposed. They are based on modelling the LSF data as a mixture of Gaussian [12], beta [13], or Dirichlet [14] distributions. The practical solutions present in recent speech codecs such as AMR-WB [15] or G.718 [16] are however based on multistage vector quantizer structures. The LPC quantizer in Opus [17] is based on a Gaussian mixture-like method, but within a variable bitrate approach.

Similarly to other current codecs, the newly 3GPP standardized EVS codec processes narrowband (NB), wideband (WB), superwideband (SWB) and fullband (FB) signals, but the entry bitrate for WB signals is 7.2kbps, for SWB is 9.6kbps and for FB it is already 16.4kbps. The input signal variety and implicitly of its spectral content prompts for having several bit allocations for the spectrum quantizer. This is needed such that it accommodates the various bitrates corresponding to the rest of the model. For instance at 13.2kbps there is one bit allocation for the LPC quantizer within the frequency range of 0-6.4kHz for WB signals and a lower one for the same range spectrum of SWB signals, because some of the bits are allocated to the higher part of the spectrum. In addition, in order to have an efficient encoding, different codebooks are used for the different internal coding modes of the codec, based on the signal type, e.g. voiced, unvoiced, transition, etc. The variability of the coding bitrates and modes requires a flexible structure that can handle multiple cases while keeping the size of the codebooks within reasonable limits.

This paper presents a method for encoding LSF's using multiple scale lattice vector quantization (MSLVQ) combined with a trellis based scheme. A bit exact lower complexity version of the MSLVQ scheme is also proposed. The overall spectrum shape encoding technique is experimentally shown to have better coding efficiency, lower complexity and lower table ROM consumption in comparison with other LSF quantization methods from state of the art codecs.

2. SPECTRUM QUANTIZATION IN THE EVS CODEC

The recently 3GPP standardized EVS speech and audio codec is mainly a core switching codec [18], with a CELP core for speech signals. In the CELP core there is an internal signal type classifier differentiating between inactive, unvoiced, voiced, generic, transition and audio type signals. Since the spectrum is different for all of these 6 signal types, it is useful to use the classifier information and design separate LSF codebooks for each coding type. In addition, the codec handles NB, WB, SWB and FB signals. Consequently, the spectrum type is further classified to be used for NB i.e. used for NB signals, for WB (spectrum up to 6.4kHz) i.e. used for WB signals processed internally at 12.8kHz, and for WB2 (spectrum up to 8kHz) i.e. used for the WB, SWB, FB signals that are processed internally at 16kHz.

2.1. Predictor type selection for LSF quantization

Due to the nature of the spectral data for each of the considered coding types, three cases are considered: one without any prediction,

denoted safety-net, a second one, purely predictive, using a first order MA predictor, and the third one, a switched safety net/predictive mode, using first order AR predictors. The predictor type selection across the coding types is presented in Table 1. For coding types with

	I	UV	V	G	T	A
NB	1	1	2	2	0	2
WB $<$ 9.6kbps	1	1	2	2	0	2
WB $\geq$ 9.6kbps	1	1	2	1	0	1
WB2	1	-	2	1	0	1

Table 1: Predictor allocation for each of the coding types: inactive (I), unvoiced (UV), voiced (V), generic (G), transition (T), audio (A). The values in the table correspond to safety net only - 0, MA prediction -1, switched safety net/AR prediction - 2. The UV mode for WB2 is not used.

high interframe correlation the AR predictor is used because in the considered cases, it brings the highest coding gain. For these coding types, to reduce propagation of errors across frames, the safety net mode is also used. For coding types with very small interframe correlation only the safety net mode is used and for the remaining coding types the MA predictor is used. The predictor values are optimized for all quantizer modes. For a given coding mode and bandwidth, all bitrates use the same predictor values. In general LSF values for voiced speech are considered quite stable over several consecutive frames. Consequently the corresponding AR predictor has the highest coefficient values. Other AR predictor coefficients are slightly lower. For the MA predictor the same value of $\frac{1}{3}$ is used everywhere. The value is significantly lower than for AR coefficients since the quantization error starts oscillating over time if the MA coefficient is too large. The value is experimentally chosen to provide reasonable prediction efficiency, stability and good error recovery.

2.2. LSF quantizer structure

This section presents the high level structure description of the LSF quantizer specifying number of stages and bits per stage for each coding mode where applicable. A safety net, predictive or switched safety-net predictive multi-stage vector quantizer (MSVQ) is used to quantize the full length LSF vector for all coding types except voiced at 16 kHz internal sampling frequency. The last stage of the MSVQ is a multiple scale lattice vector quantizer (MSLVQ). For each coding mode a number of 1 to 4 unstructured VQ stages are used, followed by the MSLVQ stage. The number of stages and number of bits per each stage for each coding mode are detailed in Table 2. They have been decided such that the table ROM consumption is not too high while keeping the quantization distortion under reasonable values. The LSF codebooks were trained with clean and noisy speech signals as well as music signals. All training data was collected from inside the codec, so that each LSF training vector was mapped to the correct mode and bandwidth before codebook training. EVS codec classifier performs slightly differently according to the requested bitrate, so all bitrates had to be used when collecting LSF codebook training vectors. Since LSF vectors represent the input signal spectrum they are totally dependent on the input signal characteristics such as bandwidth and e.g. capture device frequency response characteristics and naturally all speakers sound unique. In order to avoid over training to some specific database the training data was randomly collected from various internal and open source speech databases such as VoxForge and LibriVox [19], [20]. Music data was ripped from random CDs. In addition all signals were randomly filtered with different filters available for ITU-

Coding type	Bits SN	Bits SN stages	Bits Pred	Bits Pred stages
I NB,WB,WB2	-	-	5	5
UV NB	-	-	8	4+4
V NB, WB	8	4+4	6	3+3
G NB, WB	9	5+4	6	3+3
T NB, WB	9	5+4	-	-
A NB, WB	4	4	0	0
UV WB	-	-	12	4+4+4
G, A WB2	-	-	5	5
T WB2	8	4+4	-	-
CNG	4	4	-	-
G WB$\geq$ 9.6kbps	-	-	5	5

Table 2: Bits allocation in the VQ stages for safety net (SN) and predictive (P) modes of the LSF quantizer.

Tools filter package [21]. For some of the speech samples additional background noise excerpts were added at varying SNR levels. All this was done at all supported sampling rates of 8, 16, 32, and 48kHz. The LSF MSVQ codebooks were trained using the method described in [22]. After MSVQ codebooks were considered good enough residual LSF data was calculated and MSLVQ codebooks were trained.

Some of the mode specific unstructured codebooks whose sizes are specified in Table 2 are actually common between modes in order to save storage. The NB voiced and generic mode share the predictive mode codebooks. The NB generic and transition mode share the safety net codebooks. The same applies for the similar WB modes.

The total number of bits allocated to the LSF quantization depends on the overall bitrate. There are 17 CELP core bitrates and the bits allocated for each coding type are as follows: for inactive 22, 31, 41 bits, for unvoiced 27, 31, 37, 40 bits, for voiced 16, 31, 35, 36, 37, 38, 39 bits, for generic 22, 29, 31, 33, 36, 37, 38, 39, 41 bits, for transition 31, 33, 34, 40, 41 bits, for audio 22, 31 bits and for confort noise generation (CNG) 29 bits. Note that a specified CELP core bitrate does not necessarily correspond to the overall bitrate of the codec, but it can also correspond to an internal bitrate of the codec used in conjunction with a superwideband extension for instance. The CNG mode is used in discontinuous transmission mode.

The 16-dimensional mean removed LSF vector is quantized in the 16 dimensional space with the first VQ stages and two candidates are retained. The residual vector for each candidate is split into two 8-dimensional subvectors that are each quantized with an MSLVQ structure as described in the following section. The LSF vector for voiced coding type for internal sampling frequency of 16kHz is encoded with the block constrained trellis coded vector quantization technology detailed in section 4.

3. MULTI STAGE MULTIPLE SCALE LATTICE VECTOR QUANTIZATION (MS-MSLVQ)

The multiple scale lattice vector quantization scheme has been proposed in [11]. We will briefly describe it here and present how it has been used in order to accommodate all LSF bitrates in the EVS codec. In addition, complexity reduction of the MSLVQ encoding process is proposed.

A lattice is defined as an infinite set of points that are evenly distributed in the n-dimensional space. The lattice used in this work is D_8^+ and one of its definitions is based on the lattice D_8. The lattice

D_8 is defined as: $D_8 = \{x \in \mathbb{Z}^8 \Big| \sum_{i=0}^{8-1} x_i \text{is an even integer}\}$ and the lattice D_8^+ is defined as: $D_8^+ = D_8 \cup \Big(\big(\frac{1}{2}, ..., \frac{1}{2}\big) + D_8\Big)$. When used as a codebook, only a subset of the lattice is considered. Customary this subset is delimited by setting a maximum l_p- norm of the lattice points in it. This lattice subset is dubbed lattice truncation. If l_2-norm is used the truncation is called spherical truncation.

A leader vector of a lattice is defined as a vector with positive elements ordered in decreasing order. A leader class generated by a leader vector is the set of all possible signed permutations of the leader vector components, for which some constraints related to the signs can apply. A truncation of the lattice D_8^+ can be expressed as a union of leader classes defined by a set of corresponding leader vectors whose norms obey the constraint imposed by the truncation definition. The use of the leader classes allows to extend the definition of the truncation to a set of leader classes, without being constrained that all the leader vectors having the maximum allowed norm are in the truncation. To adapt to the data magnitude, usually a scale is associated to the lattice truncation points in order to obtain a lattice codebook.

A multiple scale lattice VQ is defined [11] as a union of several lattice truncations differently scaled. If three truncations are considered, which is generally sufficient, there are three integer numbers corresponding to the maximum number of leader vectors included in each truncation, and three floating point numbers corresponding to the scales associated to each truncation. Therefore for each bitrate allocated for the LSF vector, in addition to the codebooks from the first stages, for each of the two 8-dimensional subvectors of the LSF residual vector, an unique MSLVQ structure is defined by the 6 above mentioned numbers. These numbers are obtained using the training procedure described in [23].

The encoding of an 8-dimensional residual LSF vector is described next. We propose here an alternative encoding method to the search on leaders [11], using a transposed version of the input vector and reducing the overall average encoding complexity of the LSF vector from 3.12WMOPS to 1.18WMOPS while the maximum encoding complexity is halved.

Suppose x is the current LSF 8-dimensional subvector and w its corresponding weight vector. The vector x is normalized, i.e. component wise divided by the off line estimated standard deviation. The resulting vector is further sorted in descending order based on the absolute value of its components and the weights vector is arranged following the same order. Let x' be the vector sorted in descending order of the normalized absolute values of x and w' the correspondingly sorted weights vector. The weighted distance to the best codevector of each leader class corresponds to:

$$\begin{aligned}\|x' - s_j l_k\|^2_{w'} \quad &= \sum_{i=1}^{8} x'^2_i w'_i - 2s_j \sum_{i=1}^{8} w'_i x'_i l_{ki} \\ &\quad + s_j^2 \sum_{i=1}^{8} w'_i l^2_{ki} \end{aligned} \tag{1}$$

where l_k is the leader vector corresponding to class k and s_j is the scale of the truncation j. Each lattice codebook has at most 3 truncations with their corresponding scales. Each truncation has a given number of leader vector classes. The sum of cardinalities of the classes for the truncations forming the codebook for the first LSF subvector and for the second subvector are within the number of bits for the considered operating point given by the overall bitrate and bandwidth. Computing in the transformed input space only the second and the third terms from Equation 1 directly gives a relative measure of goodness for the best codevector from the leader class k and truncation j:

$$d_{kj} = -2s_j \sum_{i=1}^{8} w'(i) x'_i l_{ki} + s_j^2 \sum_{i=1}^{8} w'_i l^2_{ki} \tag{2}$$

The part of Equation 2 that is independent of the scale is calculated only once for all the leader classes from the first truncation.For this purpose the first truncation is chosen to have the highest number of leader classes. The contribution of the scale values is considered only later, in order to obtain the value d_{kj}. The leader class vector l_k and the truncation j having the smallest d_{kj} correspond to the codevector of the current input vector. The inverse permutation of the sorting operation on the input vector applied on the winning leader vector l_k gives the lattice codevector after applying the corresponding signs with the parity constraint [11]. The final codevector is obtained after multiplication with the scale s_j and with the inverse of the component-wise off-line computed standard deviation. The standard deviations are individually estimated for each coding mode and bandwidth.

The candidate quantized LSF vectors are obtained by adding each lattice quantized residual to the corresponding candidates from the upper stages. The obtained candidates are sorted in increasing order. For each sorted candidate the weighted Euclidean distortion with respect to the original LSF vector is calculated. The candidate that minimizes this distortion is selected as codevector to be encoded. The indexes corresponding to the first unstructured optimized VQ codebooks together with the index in the lattice codebook are written in the bitstream. The lattice codevectors indexes are obtained using binomial enumeration [24].

For the coding types that have both predictive and safety net modes one of the modes is selected based on the weighted Euclidean distortion. To increase the resilience to frame erasure errors, preference is given to the safety-net if its corresponding distortion is small enough. For internal sampling rate of 16kHz, the decision between predictive and safety-net mode is done in open loop to reduce complexity (see section 4).

4. BLOCK-CONSTRAINED TRELLIS CODED VECTOR QUANTIZATION

The voiced coding (VC) type of EVS at 16 kHz internal sampling frequency has two encoding rates, 31 and 40 bits per frame, as well as two encoding schemes, predictive and safety-net. In the VC type, the LSF is quantized by a 16-state, 8-stage block-constrained trellis coded vector quantization (BC-TCVQ) scheme. This is a low-complexity quantization scheme requiring exactly one bit per source sample to specify the trellis path [25], [26]. The quantization scheme for VC type is selected in an open-loop manner as illustrated in Figure 1, where the open-loop selection operates at half the computational complexity compared with selection using closed-loop algorithms. In the figure, the prediction error E_k of the k-th frame is obtained as

$$E_k = \sum_{i=0}^{M-1} w(i)\big(z_k(i) - p_k(i)\big)^2 \tag{3}$$

where $p_k(i) = \rho(i)\hat{z}_{k-1}(i)$, for $i = 0, ..., M-1$. $\rho(i)$ are the selected Auto Regressive (AR) prediction coefficients, $\hat{z}_{k-1}(i)$ is the mean-removed quantized LSF of the previous frame, and M is the LPC order. When E_k is greater than a threshold, the current frame is considered to be non-stationary, in which case the safety-net scheme is chosen. Otherwise the predictive scheme is selected.

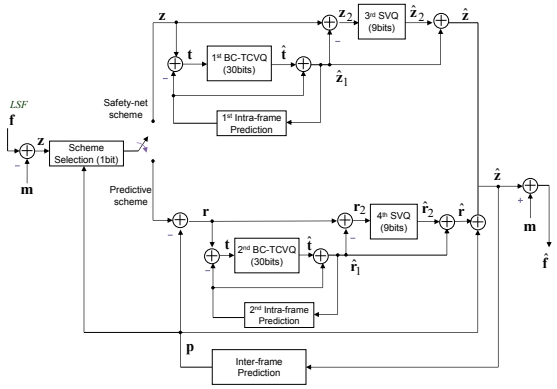

Fig. 1: Block diagram of BC-TCVQ/SVQ with safety-net for encoding rate of 31 and 40 bits/frame.

In EVS, the encoding scheme selection, 1st and 2nd BC-TCVQ, and 1st and 2nd intra-frame prediction blocks of the 40-bit LSFQ are exactly the same as those of the 31-bit LSFQ, to minimize memory size. For the 31-bit LSFQ, if the safety-net scheme is selected, the mean-removed LSF vector, $z_k(i)$, is quantized by the 1st BC-TCVQ and 1st intra-frame prediction with 30 bits. If the predictive scheme is selected, the prediction error, $r_k(i)$, is quantized by the 2nd BC-TCVQ and 2nd intra-frame prediction with 30 bits. Intra-frame correlation is typically present in the inter-frame AR prediction error vectors. The intra-frame prediction uses the quantized elements of the previous stage. The prediction coefficients used for the intra-frame prediction are predefined by the codebook training process. The prediction coefficients are two-by-two matrices for the 2-dimensional vector. The intra-frame prediction process of BC-TCVQ is as follows. The prediction residual vector, $t_k(i)$ which is the input of the 1st BC-TCVQ, is computed as

$$t_k(0) = z_k(0)$$

$$t_k(i) = z_k(i) - \tilde{z}_k(i), \text{for } i = 1, ..., \frac{M}{2} - 1 \tag{4}$$

where $\tilde{z}_k(i) = \mathrm{A}_i \hat{z}_k(i-1), \text{for } i = 1, ..., \frac{M}{2} - 1$ and $\tilde{z}_k(i)$ is the estimation of $z_k(i)$, $\hat{z}_k(i-1)$ is the quantized vector of $z_k(i-1)$, and A_i is the prediction matrix with 2×2 which is computed as:

$$\mathrm{A}_i = \mathrm{R}_{01}^i [\mathrm{R}_{11}]^{-1}, \text{for } i = 1, ..., \frac{M}{2}, \tag{5}$$

where $\mathrm{R}_{01}^i = E[\mathrm{z}_i \mathrm{z}_{i-1}^t]$ and $\mathrm{R}_{11}^i = E[\mathrm{z}_{i-1} \mathrm{z}_{i-1}^t]$ and M is the LPC order. Then

$$\hat{z}_k(i) = \hat{t}_k(i) + \tilde{z}_k(i), \text{for } i = 0, ..., \frac{M}{2} - 1. \tag{6}$$

The prediction residual, $t_k(i)$, is quantized by the 1st BC-TCVQ. The 1st BC-TCVQ and the 1st intra-frame prediction are repeated to quantize $z_k(i)$.

In the LSF quantization with 40 bits per frame, the difference between the mean-removed LSF and its BC-TCVQ output is quantized by the 3rd and 4th Split VQ (SVQ) with 9 bits. The 3rd SVQ is exactly the same as the 4th SVQ, as both SVQs use an identical codebook. Since the input distribution of the 3rd SVQ is different from that of the 4th SVQ, scaling factors computed with the distribution of the residual signals z_2 and r_2 are used to compensate for the difference. Table 3 shows the bit allocation for LSFQ at 30 and 40 bits/frame.

Parameter		Bit allocation
Scheme selection		1
BC-TCVQ	Path information (Initial states + path + final states)	2+4+2
	Subset codewords	4 bits×2 (Stages 1 to 2) 3 bits ×2 (Stages 3 to 4) 2 bits ×4 (Stages 5 to 8)
SVQ	Subset codevectors	5 (1st vector with dim.=8) 4 (2nd vector with dim.=8)
Total		31/40

Table 3: Bit allocation for LSFQ at 31 and 40 bits/frame

5. RESULTS AND DISCUSSION

The LSF encoding obtained by the combination of the two previously presented methods forms the LPC quantization block in the EVS codec. Experiments are performed on combined speech, noisy speech, and music signals for different bandwidths. The results in terms of average spectral distortion (SD) and spectral distortion outlier distribution are presented in Table 4 and compared with a multistage VQ structure similar to the one in G.718. Exact comparison of the EVS and G.718 codecs, or any other codec, is difficult to perform because they have different coding mode classifiers and different bitrates. For the ease of comparison, only the case of 31 bits has been used for all modes. The multistage approach MSVQ has 5 stages of 7,6,6,6,6 bits respectively. It uses MA predictor for all modes, the same as the one used in the MS-MSLVQ case. The table ROM is reduced from 36.9kB for the multistage VQ to 24.6kB for multistage MSLVQ and BC-TCVQ combined. The BC-TCVQ structure takes by itself 2.432 kB. The average encoding complexity decreases from 4.422 WMOPS for the multistage structure to 1.951 WMOPS for the BC-TCVQ and to 1.189 WMOPS for the multi-stage MSLVQ structure. The decoding average complexity is negligible in all cases, below 0.1 WMOPS.

	MS-MSLVQ+BC-TCVQ			MSVQ		
Coding type	Av. SD (dB)	[2,4] (%)	>4 (%)	Av. SD (dB)	[2,4] (%)	>4 (%)
I NB	**1.856**	**34.07**	1.19	1.976	46.74	1.11
UV NB	**1.466**	**12.61**	**0.37**	2.931	90.83	5.36
V NB	**1.084**	**2.63**	0.18	1.507	18.92	0.05
GE NB	**1.281**	**5.92**	**0.18**	1.999	46.23	0.75
Total NB	**1.243**	**6.07**	**0.22**	1.928	41.54	1.01
I WB	**1.488**	**17.37**	**0.60**	1.665	24.55	1.60
UV WB	**1.587**	**13.28**	**0.82**	1.675	20.43	1.65
V WB	**1.431**	**11.94**	**0.33**	1.615	16.96	0.16
GE WB	1.677	21.00	0.82	1.673	23.35	0.78
Total WB	**1.583**	**17.07**	**0.65**	1.653	20.88	0.67
I WB2	**1.567**	**21.72**	**0.17**	1.810	34.66	1.38
UV WB2	**1.537**	**16.08**	**0.78**	2.945	91.08	18.43
V WB2	**1.440**	**13.93**	**0.18**	1.773	28.85	0.23
GE WB2	**1.789**	**26.95**	**0.87**	1.871	37.03	1.18
A WB2	**1.657**	**22.36**	2.93	1.693	27.18	0.66
Total	**1.664**	**22.36**	**1.30**	1.807	20.88	0.67

Table 4: Average SD and SD outliers distribution (percentage of frames having SD between 2 and 4 dB, ([2,4]), and percentage of frames having SD larger than 4dB, (>4)) for MS-MSLVQ with BC-TCVQ (V WB2) and MSVQ at 31 bits.

6. REFERENCES

[1] F.K. Soong and B.-H. Juang, "Line spectrum pairs (LSP) and speech data compression," in *Proceedings of ICASSP*, 1984, pp. 1.10.1–1.10.4.

[2] W.R. Gardner and B.D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. on Speech and Audio Processing*, vol. 3, pp. 367–381, September 1995.

[3] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. on Speech and Audio Processing*, vol. 1, pp. 3–14, January 1993.

[4] J.R.B. De Marca, "An LSF quantizer for the North American half-rate speech coder," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, pp. 413–419, 1994.

[5] W.P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4kb/s speech coding," *IEEE Trans. on Speech and Audio Processing*, vol. 1, pp. 373–385, October 1993.

[6] Y. Tanaka and T. Taniguchi, "Efficient coding of LPC parameters using adaptive prefiltering and MSVQ with partially adaptive codebooks," in *Proceedings of ICASSP*, 1993, pp. 5–8.

[7] N. Phamdo, N. Farvardin, and T. Moriya, "A unified approach to tree-structured and multi-stage vector quantization for noisy channels," *IEEE Transactions on Information Technology*, vol. 39, no. 3, pp. 835–850, 1993.

[8] M. Xie and J.-P. Adoul, "Algebraic vector quantization of LSF parameters with low storage and complexity computation," *IEEE Trans. on Speech and Audio Processing*, vol. 4, pp. 234–239, May 1996.

[9] J. Pan, "Extension of two-stage vector quantization-lattice vector quantization," *IEEE Trans. on Communications*, vol. 45, pp. 1538–1547, December 1997.

[10] J. Pan and T.R. Fischer, "Vector quantization of speech line spectrum pair parameters and reflection coefficients," *IEEE Trans. on Speech and Audio Processing*, vol. 6, pp. 106–115, March 1998.

[11] A. Vasilache, B. Dumitrescu, and I. Tăbuş, "Multiple-scale leader-lattice VQ with application to LSF quantization," *Signal Processing*, vol. 82, pp. 47–70, 2002.

[12] A.D. Subramaniam and B.D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. on Speech and Audio Processing*, vol. 11, pp. 130–142, March 2003.

[13] Z. Ma and A. Leijon, "PDF-optimized LSF vector quantization based on mixture models," in *Proceedings of INTERSPEECH*, 2010, pp. 2374–2377.

[14] Z. Ma, A. Leijon, and W. B. Kleijn, "Vector quantization of LSF parameters with a mixture of Dirichlet distributions," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 21, pp. 1777–1790, September 2013.

[15] 3GPP, "Technical specification group service and system aspects; audio codec processing functions; extended AMR wideband codec; transcoding functions," in *TS26.171*.

[16] ITU-T, "G.718 frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s," 2008.

[17] K. Vos, K.V. Sørensen, S.S. Jensen, and J.-M. Valin, "Voice coding with Opus," in *Proceedings of* 135^{th} *AES Convention*, 2013.

[18] M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, H. Pobloth, L. Miao, Z. Wang, L. Laaksonen, A. Vasilache, Y. Kamamoto, K. Kikuiri, S. Ragot, J. Faure, H. Ehara, V. Rajendran, A. Venkatraman, H. Sung, E. Oh, H. Yuan, and C. Zhu, "Overview of the EVS codec architecture," in *Proceedings of ICASSP 2015*, 2015.

[19] VoxForge, "Voxforge open source speech database," Sept. 2014, online: http://www.voxforge.org/.

[20] LibriVox, "Free public domain audiobooks," Sept. 2014, online: https://librivox.org/.

[21] ITU-T G.191, *Software tools for speech and audio coding standardization*, Mar. 2010.

[22] W.P. LeBlanc, B. Bhattacharya, S.A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 373–385, 1993.

[23] A. Vasilache and I. Tăbuş, "Image coding using multiple scale leader lattice vector quantization," in *Proceedings of SPIE 5014, Image Processing: Algorithms and Systems II*, 2003.

[24] A. Vasilache and I. Tăbuş, "Robust indexing for lattices and permutation codes over binary symmetric channels," *Signal Processing*, vol. 83, pp. 1467–1486, 2003.

[25] S. Kang, Y. Shin, and T.R. Fischer, "Low-complexity predictive trellis-coded quantization of speech line spectral frequencies," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2070–2079, July 2004.

[26] J. Park and S. Kang, "Block constrained trellis coded vector quantization of LSF parameters for wideband speech codecs," *ETRI Journal*, vol. 30, pp. 738–740, October 2008.