Interspeech 2018
2-6 September 2018, Hyderabad

Multi-frame Coding of LSF Parameters using Block-Constrained Trellis Coded Vector Quantization

Yaxing Li[1], *Shan Xu*[1], *Shengwu Xiong*[1,*], *Anna Zhu*[1], *Pengfei Duan*[1], *Yueming Ding*[2]

[1]School of Computer Science and Technology, Wuhan University of Technology, China
[2]School of Computer Science and Engineering, Tianjin University of Technology, China
yaxing.li@whut.edu.cn, xiongsw@whut.edu.cn

Abstract

In this paper, the predictive block-constrained trellis coded vector quantization (BC-TCVQ) schemes are developed for quantization of multiple frames line spectral frequency (LSF) parameters. The consecutive LSF frames are interleaved to subvectors for trellis modeling. The predictive BC-TCVQ systems are then designed to encode multi-frame LSF parameters. The performance evaluation of proposed schemes is compared with the single-frame LSF encoding methods using multi-stage vector quantization (MSVQ) and predictive block-constrained trellis coded quantization (BC-TCQ), demonstrating significant reduction of bit rate for transparent coding. The developed multi-frame LSF quantization schemes show satisfactory performance even at very low encoding rate, and thus can be efficiently applied to the speech coders with moderate delay.

Index Terms: Line spectral frequency (LSF), quantization, speech coding, block-constrained trellis coded vector quantization (BC-TCVQ)

1. Introduction

In the linear predictive coding (LPC) based speech coders, the encoding of line spectral frequency (LSF) parameters is a major issue [1]. LSF parameters are usually extracted, quantized, and transmitted on a frame-by-frame basis in speech coders. The empirical lower bounds on the bit rate required for transparent coding of single frame LSF parameters have been studied in [2–5]. However, high correlation between consecutive LPC parameters has been evidenced and it can be exploited to surpass the bit rate limits of single-frame LSF quantization [6, 7]. Paliwal and So introduced the multi-frame GMM-based block quantizer for quantizing the LSF parameters for narrowband and wideband speech coding [6, 7]. Recently, a multi-frame quantization of LSF parameters using a deep autoencoder and pyramid vector quantizer was proposed [8]. The multi-frame coding of LSF parameters also has been extensively applied to the speech coders with moderate delay to reduce the encoding rate. For the adaptive multi rate (AMR) coder at 12.2 kbps mode, the analyzed two sets of LPC coefficients are jointly quantized using split matrix quantization [9]. For the enhanced mixed excitation linear prediction (MELPe) coder at 1200 bps and 600 bps mode, the respective three and four consecutive LSF frames are quantized using the multi stage vector quantization (MSVQ) [10]. The analyzed multiple LSF frames are also jointly encoded in the recently developed variations of the MELPe coder [11–13].

Marcellin and Fisher introduced a trellis coded quantization (TCQ) to encode the LSF parameters [14]. TCQ achieved better performance than other conventional quantization methods in terms of distortion and complexity. Trellis coded VQ (TCVQ) generalizes TCQ to allow vector codebooks and branch labels for further performance improvement [15]. In the traditional TCQ and TCVQ, an initial trellis state is encoded as side information which is an additional rate for source samples or vectors. Block-constrained TCQ (BC-TCQ) requires exactly one bit per source sample to specify the trellis path with low complexity [16]. Block-constrained TCVQ (BC-TCVQ) of LSF parameters for wideband speech coders was also proposed by combining TCVQ with BC-TCQ [17–19].

The existing trellis encoding methods for LSF parameters [16–19] are single-frame quantization schemes. In this paper, the efficient BC-TCVQ is developed for multi-frame LSF parameters quantization. The successive K frames LSF vectors are firstly interleaved to subvectors for trellis modeling. The number of interleaved subvectors equals to the order of LSF parameters and the elements in each subvector correspond to the same dimensional components of the K frames LSF vectors. The dependencies of the interleaved subvectors are modeled by using the block-constrained trellis structure, in which each trellis stage is associated with one interleaved subvector. The predictive BC-TCVQ systems are then designed to encode multiple frames of LSF parameters.

2. Multi-frame coding of LSF parameters using BC-TCVQ

2.1. Trellis modeling of multi-frame LSF parameters

The discussed trellis structures are based on a rate-1/2 convolutional code, which has $N=2^v$ trellis states and two branches entering/leaving each trellis state [14–19]. For any $0 \le s \le v$, a block-constrained trellis structure that allows 2^s initial trellis states and exactly 2^{v-s} terminal trellis states for each allowed initial trellis state. Given a block of m source scalars or vectors, a single Viterbi algorithm encoding [20], starting from the allowed initial trellis states, proceeds in the normal trellis search up to the stage m-s. A unique terminating path, possibly dependent on the initial trellis state, is pre-specified for each trellis state at stage m-s through stage m.

For the BC-TCQ of LSF parameters, each trellis stage corresponds to one dimension element of the LSF vector [16]. As to the BC-TCVQ of LSF parameters for wideband speech, one trellis stage is associated with two neighboring elements of the LSF parameters [17–19]. The LSF parameters show the natural ordering property in the same frame which indicates their intra-frame correlation dependencies [1, 21]. The associated one dimension element or two neighboring elements in each stage maintain the ordering property, which

 10.21437/Interspeech.2018-2578

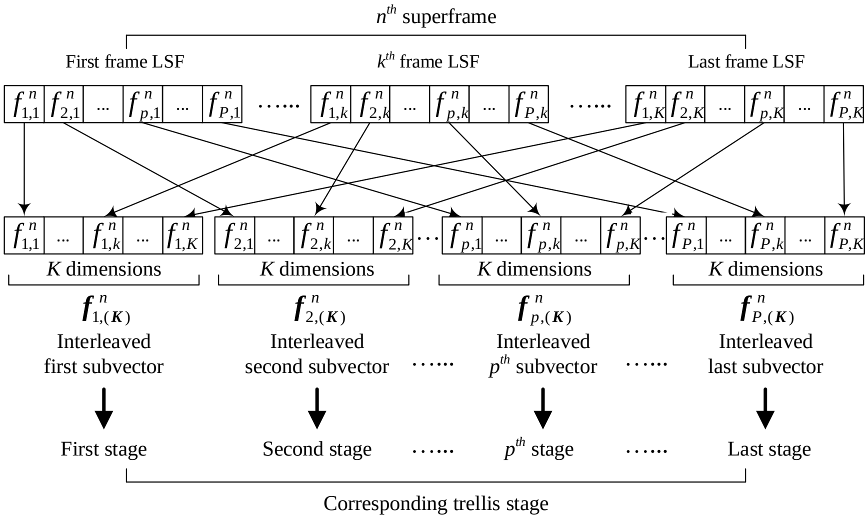

Figure 1: *Interleaved subvectors and associated stages for trellis modeling.*

can be exploited to design the intra-frame predictive BC-TCQ and BC-TCVQ, respectively [16–19].

For the multi-frame LSF parameters quantization, K successive speech frames are grouped into a superframe. The LSF parameters of each frame are obtained using a P^{th} order LPC analysis method applied on the short-term analysis window. The k^{th} frame LSF vector for the superframe index n is denoted as $[f^n_{1,k}, f^n_{2,k} \cdots f^n_{P,k}]^T$, for k=1 to K. The concatenated K frames LSF parameters need to be split into a proper number of subvectors for trellis modeling. The boundary elements of two adjacent LSF vectors violate the ordering property, which restricts the design of the predictive quantizer. Thus, the K successive frame LSF vectors in a superframe are interleaved to P subvectors, in which the same dimensional parameters of each LSF vector are grouped together. The number of elements in an interleaved subvector is same and equal to the superframe size K. Figure 1 shows the interleaved subvectors and associated stages for trellis modeling. The interleaved subvectors are denoted as $\boldsymbol{f}^n_{(\boldsymbol{P}),(\boldsymbol{K})}$ and the p^{th} subvector is represented as $\boldsymbol{f}^n_{p,(\boldsymbol{K})} = [f^n_{p,1}, f^n_{p,2} \cdots f^n_{p,K}]^T$, for p=1 to P. P is equal to 10 for narrowband telephone speech.

In this paper, the BC-TCVQ for multi-frame LSF parameters possesses a trellis structure with 16 states (N=16, v=4) and 10 stages (P=10). The interleaved 10 LSF subvectors are allocated to each trellis stage. Four initial and terminal states (s=2) are used in the block-constrained structure, i.e., two bits are used for both the initial state and terminal state [16–19]. The ordering property is still maintained in the interleaved subvectors, which can be exploited to design the predictive BC-TCVQ for encoding multi-frame LSF parameters.

2.2. Predictive BC-TCVQ for multi-frame LSF parameters coding

There are significant dependencies between elements within a LSF vector and between consecutive LSF vectors, which are commonly referred to as intra-frame correlation and inter-frame correlation, respectively. The highest quantization efficiency can be achieved when both intra-frame and inter-frame correlations are removed. Since the subvectors are interleaved across frames, the subvector-based intra-superframe prediction exploits both the intra-frame and inter-frame correlations of the consecutive LSF vectors in a superframe. An intra-superframe predictive quantization scheme for the interleaved K consecutive LSF vectors $\boldsymbol{f}^n_{(\boldsymbol{P}),(\boldsymbol{K})}$ with a first order autoregressive (AR) predictor is shown in Figure 2.

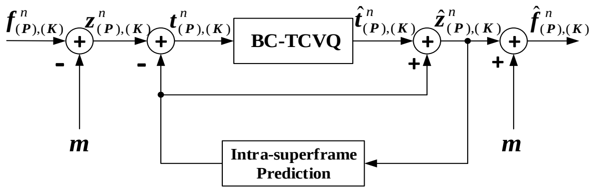

Figure 2: *Predictive BC-TCVQ with first order AR intra-superframe predictor.*

The prediction residual, $\boldsymbol{t}^n_{(\boldsymbol{P}),(\boldsymbol{K})}$ which is the input of the BC-TCVQ, is computed as

$$\boldsymbol{t}^n_{1,(\boldsymbol{K})} = \boldsymbol{z}^n_{1,(\boldsymbol{K})}, \tag{1}$$

$$\boldsymbol{t}^n_{p,(\boldsymbol{K})} = \boldsymbol{z}^n_{p,(\boldsymbol{K})} - \tilde{\boldsymbol{z}}^n_{p,(\boldsymbol{K})}, \quad 1 < p \le P, \tag{2}$$

where $\tilde{\boldsymbol{z}}^n_{p,(\boldsymbol{K})}$ is the estimation of $\boldsymbol{z}^n_{p,(\boldsymbol{K})}$, which is the mean-removed version of $\boldsymbol{f}^n_{p,(\boldsymbol{K})}$. $\tilde{\boldsymbol{z}}^n_{p,(\boldsymbol{K})}$ is computed as

$$\tilde{\boldsymbol{z}}^n_{p,(\boldsymbol{K})} = \boldsymbol{A}_{p,(\boldsymbol{K})} \hat{\boldsymbol{z}}^n_{p-1,(\boldsymbol{K})}, \quad 1 < p \le P, \tag{3}$$

where $\hat{\boldsymbol{z}}^n_{p-1,(\boldsymbol{K})}$ is the quantized vector of $\boldsymbol{z}^n_{p-1,(\boldsymbol{K})}$, and $\boldsymbol{A}_{p,(\boldsymbol{K})}$ is the K by K intra-superframe prediction matrix. For simplicity, the prediction matrix is calculated by the open-loop approach using the training data [18], and is given by

$$\boldsymbol{A}_{p,(\boldsymbol{K})} = \boldsymbol{R}^p_{01} \left[\boldsymbol{R}^p_{11} \right]^{-1}, \quad 1 < p \le P, \tag{4}$$

where $\boldsymbol{R}^p_{01}$ and $\boldsymbol{R}^p_{11}$ are both K by K matrix and are given as

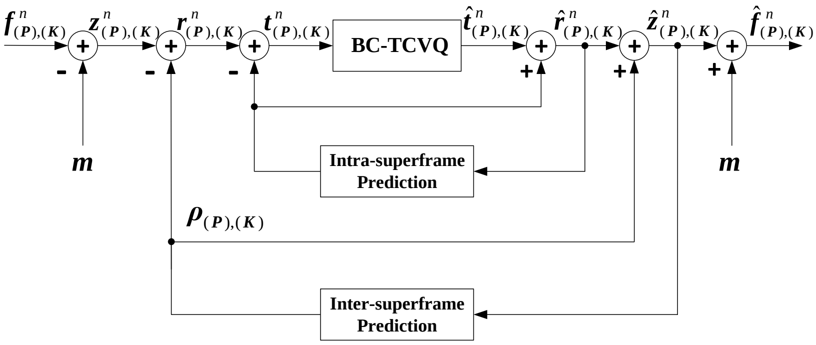

Figure 3: *Predictive BC-TCVQ with first order AR intra-superframe predictor and first order AR inter-superframe predictor*.

$$\boldsymbol{R}_{01}^{p} = E[\boldsymbol{z}_{p,(\boldsymbol{K})}^{n}(\boldsymbol{z}_{p-1,(\boldsymbol{K})}^{n})^{T}], \tag{5}$$

$$\boldsymbol{R}_{11}^{p} = E[\boldsymbol{z}_{p-1,(\boldsymbol{K})}^{n}(\boldsymbol{z}_{p-1,(\boldsymbol{K})}^{n})^{T}]. \tag{6}$$

When the mean-removed subvector $\boldsymbol{z}_{p,(\boldsymbol{K})}^{n}$ is to be quantized, its prediction is formed using $\hat{\boldsymbol{z}}_{p-1,(\boldsymbol{K})}^{n}$ which is the quantized version of its previous subvector $\boldsymbol{z}_{p-1,(\boldsymbol{K})}^{n}$. The difference of $\boldsymbol{z}_{p,(\boldsymbol{K})}^{n}$ and its prediction is then quantized via BC-TCVQ.

There is dependence of the LSF parameters in current superframe and the last frame LSF parameters in previous superframe. An inter-superframe and intra-superframe predictive BC-TCVQ is depicted in Figure 3. The AR inter-predictor general perform better than moving average (MA) method in the error free transmission condition [22]. For the AR predictor, there is no noticeable performance gain after using predictor order higher than one. In this paper, a first-order AR predictor is used to exploit inter-superframe dependence, as shown in Figure 3. At the p^{th} stage, the inter-superframe prediction is based on $\hat{z}_{p,K}^{n-1}$, which is the quantized mean-removed component of the K^{th} frame LSF vector in previous superframe. Let $\boldsymbol{r}_{p,(\boldsymbol{K})}^{n}$ be the inter-superframe prediction error vector at the p^{th} stage for the superframe index *n*. The relationship between mean-removed subvector $\boldsymbol{z}_{p,(\boldsymbol{K})}^{n}$ and prediction error vector $\boldsymbol{r}_{p,(\boldsymbol{K})}^{n}$ is modeled as

$$\boldsymbol{r}_{p,(\boldsymbol{K})}^{n} = \boldsymbol{z}_{p,(\boldsymbol{K})}^{n} - \boldsymbol{\rho}_{p,(\boldsymbol{K})}\hat{z}_{p,K}^{n-1}, \quad 1 \le p \le P, \tag{7}$$

where $\boldsymbol{\rho}_{p,(\boldsymbol{K})}$ is inter-superframe prediction vector at the p^{th} stage and its k^{th} component $\rho_{p,k}$ is computed as

$$\rho_{p,k} = \frac{\sum_{n=1}^{N_T-1} z_{p,k}^{n+1} z_{p,K}^{n}}{\sum_{n=1}^{N_T-1} (z_{p,K}^{n})^2}, \quad 1 \le k \le K, \quad 1 \le p \le P, \tag{8}$$

where N_T is the superframe number of training data. As shown in Figure 3, the intra-superframe prediction is performed after the inter-superframe prediction and the final prediction error vector is then quantized using BC-TCVQ.

3. Performance evaluation

Experiments are conducted using the TIMIT database [23] and all of the 4200 speech utterances are down sampled to 8 kHz. A total of 3800 sentences are selected for training and the remaining 400 sentences are used for testing. This results in a connection of 196.29 minutes training speech and 20.41 minutes testing speech. The 10th order LSF vectors are calculated for every 10 ms frame with a 30 ms Hamming window. *K* successive LSF frames are concatenated together for performance evaluation. In order to exploit the human auditory characteristics, the weighted Euclidian distance in [24] is used to design the predictive BC-TCVQ for multi-frame LSF parameters. To measure the performance of the quantized LSF parameters, we employ the spectral distance (SD) [25–27], which is defined as follows:

$$SD = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left[10\log_{10} P(\omega) - 10\log_{10}\hat{P}(\omega)\right]^2 d\omega}, \quad dB \tag{9}$$

where $P(\omega)$ and $\hat{P}(\omega)$ are the LPC power spectra of the original and quantized LSF vectors, respectively.

A frame-by-frame quantization method using the MSVQ [10, 25] is used for performance comparison. The design of MSVQ and the selection of parameters follow the experimental setup in [25]. The two-stages MSVQ is made by applying the LBG algorithm [28] on the training corpus using the weighted Euclidian distance [24]. The bits are divided equally between two stages, and for odd rates, the first stage is given as an extra bit. The predictive BC-TCQ with first order AR intra-frame predictor and AR inter-frame predictor [16] is another frame-by-frame quantization scheme for comparison. The bit allocation for predictive BC-TCQ of single-frame LSF parameters and predictive BC-TCVQ of two consecutive frames LSF parameters (at an encoding rate of 20 bits/frame) are shown in Table 1. Regardless of the value of *K*, the block-constrained trellis structure requires one bit per stage, i.e., in total 10 bits, to specify the trellis path. As the *K* value increases, more bits will be allocated to specify the subset codeword. For the BC-TCQ, the minimum encoding rate is 20 bits/frame since at least 1 bit is allocated for the subset codeword of each stage, i.e., totally 10 bits for subset codeword.

In the experiment, we consider transparent quality to be achieved when the average SD is approximately 1 dB, and the 2 dB outlier frames are less than 2% [21, 29]. When this condition is valid the fraction of 4 dB outliers is negligible or

Table 1: *Bit allocation for predictive BC-TCQ of single-frame LSF and predictive BC-TCVQ of two consecutive frames LSF (at an encoding rate of 20 bits/frame).*

Parameters	Bit allocation	
	Predictive BC-TCQ	Predictive BC-TCVQ, K=2
Path information (Initial states + Path + Final states)	2+6+2	2+6+2
Subset codeword	1×10 (stages 1 to 10)	3×10 (stages 1 to 10)
Total	20 bits	20×2=40 bits

zero. Table 2 shows the simulation results for the MSVQ and predictive BC-TCQ of single-frame LSF parameters at different bit rates. The MSVQ and BC-TCQ show transparent quality at an encoding rate of 24 bits/frame and 23 bits/frame, respectively.

Table 3 depicts the results of multi-frame LSF parameters encoding using the developed predictive BC-TCVQ schemes at different bit rates. For quantizing two frames jointly (K=2), it can be observed that predictive BC-TCVQ with only intra-superframe predictor, labeled as Scheme-intra, achieves the transparent quality using 20 bits/frame. The predictive BC-TCVQ with both inter-superframe and intra-superframe predictors, denoted as Scheme-inter&intra, achieves the transparency at a low encoding rate of 18 bits/frame. Comparing the predictive BC-TCQ of single frame LSF parameters, Scheme-intra and Scheme-inter&intra with regard to two frames LSF parameters quantization can save 3 bits and 5 bits, respectively. The performance improvement of developed schemes are attributed to the bit saving of trellis path and the exploitation of correlation between consecutive LSF frames. As more frames are concatenated (K=3, K=4, K=5), the developed schemes achieve the transparent coding at a more reduced bit rate. For jointly quantizing five frames LSF parameters (K=5), the transparent coding of Scheme-intra and Scheme-inter&intra are achieved as low as 16 bits/frame and 15 bits/frame encoding rate, respectively. The overall dependence of the K^{th} frame LSF parameters in previous superframe and the current superframe diminishes as the increase of K, and hence, it narrows the performance difference between Scheme-inter&intra and Scheme-intra. Comparing with the single-frame LSF parameters quantization method, the developed multi-frame encoding schemes reduce the bit rate significantly for transparent coding.

The developed multi-frame LSF quantization schemes produce very good performance when the encoding rate is as low as 14 bits per frame or even lower bit rate for large consecutive frame number K. Three or four frames (K=3, 4)

Table 2: *Average SD and 2 dB outliers for MSVQ and predictive BC-TCQ of single-frame LSF parameters.*

Bits /frame	MSVQ		Predictive BC-TCQ	
	Avg. SD (dB)	Outliers >2 dB (%)	Avg. SD (dB)	Outliers >2 dB (%)
21	1.260	4.21	0.890	2.73
22	1.187	2.85	0.845	2.27
23	1.110	1.88	**0.814**	**1.94**
24	**1.046**	**1.25**	0.774	1.68

LSF parameters are usually encoded jointly for the speech coders with moderate delay, in which the developed multi-frame quantization schemes can be applied efficiently. In most cases, the Scheme-inter&intra produces better performance than the Scheme-intra. There are occasions, however, small inter-superframe correlation is present, when rapid changes in LSF traces are evident, and hence, the Scheme-intra is preferred. The Scheme-inter&intra can be combined with a safety-net framework [30] to provide better robustness against the outliers.

Table 3: *Average SD and 2 dB outliers of multi-frame LSF parameters encoding using developed schemes.*

K	Bits /frame	Scheme-intra		Scheme-inter&intra	
		Avg. SD (dB)	Outliers >2 dB (%)	Avg. SD (dB)	Outliers >2 dB (%)
2	16	1.372	8.52	1.029	3.81
	18	1.205	4.18	**0.902**	**1.99**
	19	1.123	2.52	0.839	1.39
	20	**1.049**	**1.59**	0.780	0.96
	21	1.001	1.22	0.739	0.71
3	16	1.158	3.64	0.945	2.04
	17	1.091	2.55	**0.888**	**1.43**
	18	**1.030**	**1.78**	0.838	0.98
	19	0.957	1.01	0.779	0.61
4	14	1.215	5.07	1.038	3.07
	16	1.069	2.26	**0.907**	**1.39**
	17	**0.989**	**1.33**	0.818	0.82
	18	0.873	0.87	0.794	0.57
5	14	1.141	3.59	1.003	2.44
	15	1.072	2.33	**0.940**	**1.54**
	16	**0.996**	**1.40**	0.877	0.95
	17	0.939	0.90	0.824	0.59

4. Conclusions

In this paper, the predictive block-constrained trellis coded vector quantization (BC-TCVQ) are developed for encoding multi-frame line spectral frequency (LSF) parameters. In the developed schemes, the bit saving of trellis path and the exploitation of correlation between consecutive LSF frames lead to the significant improvement in the quantization performance. The performance evaluation is compared with the single-frame LSF quantization method using multi-stage vector quantization (MSVQ) and predictive block-constrained trellis coded quantization (BC-TCQ), demonstrating significant reduction of bit rate for transparent coding. The developed multi-frame LSF quantization schemes show satisfactory performance even at very low encoding rate, and thus can be applied to the speech coders with moderate delay.

5. Acknowledgements

The authors acknowledge the contribution of Prof. Kang Sangwon for this work and it is with sadness to note that Prof. Kang Sangwon passed away on May 4, 2018.

This work was partially supported by National Key R&D Program of China (No. 2017YFB1402203, No. 2016YFD0101903), and funded by the Fundamental Research Funds for the Central Universities (WUT: 2018IVA028), and supported by Tianjin Major Scientific and Technological Research Plan and Program under Grant No. 16ZXHLSF00160.

6. References

[1] A. M. Kondoz, *Digital Speech: Coding for low bit rate communication systems, 2nd Edition*, Chichester, West Sussex, England: John Wiley & Sons Ltd, 2004.

[2] Z. Ma, S. Chatterjee, W. B. Kleijn, et al., "Dirichlet mixture modeling to estimate an empirical lower bound for LSF quantization," *Signal Processing*, vol. 104, no. 6, pp. 291-295, 2014.

[3] S. So, K. K. Paliwal, "Empirical Lower Bound on the Bitrate for the Transparent Memoryless Coding of Wideband LPC Parameters," *IEEE Signal Processing Letters*, vol. 13, no. 9, pp. 569–572, 2006.

[4] S. Chatterjee, T. V. Sreenivas, "Predicting VQ Performance Bound for LSF Coding," *IEEE Signal Processing Letters*, vol. 15, pp. 166–169, 2008.

[5] P. Hedelin, J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Transactions on Speech & Audio Processing*, vol. 8, no. 4, pp. 385–401, 2000.

[6] K. K. Paliwal, S. So, "Multiple frame block quantisation of line spectral frequencies using Gaussian mixture models," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, May. 2004, pp. I-149–I-152.

[7] S. So, K. K. Paliwal, "Multi-frame GMM-based block quantisation of line spectral frequencies for wideband speech coding," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Mar. 2006, pp. I-121–I-124.

[8] Y. Li, E. D. Emiru, S. Xiong, A. Zhu, P. Duan and Y. Li, "Multi-frame quantization of LSF parameters using a deep autoencoder and pyramid vector quantizer," *in Proc. INTERSPECCH* (to appear), 2018.

[9] AMR Wideband Speech Codec; Transcoding Functions (Release 5), 3GPP TS 26.190 (V5.1.0.), Dec. 2001.

[10] The 600 bit/s, 1200 bit/s and 2400 bit/s NATO interoperable narrow band voice coder, STANAG 4591, 2008.

[11] L. Zhu and Q. Li, "A 600bps Vocoder Algorithm Based on MELP," *in 2017 2nd International Conference on Electrical and Electronics: Techniques and Applications*, 2017, pp. 306–311.

[12] X. Ma, Y. Li, J. Jiang, et al, "400bps High-Quality Speech Coding Algorithm," *in 2016 International Symposium on Computer, Consumer and Control*, 2016, pp. 256–259.

[13] Y. Li, Q. Hao, P. Zhang, et al, "A variable-bit-rate speech coding algorithm based on enhanced mixed excitation linear prediction," *in 2016 9th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics*, 2016, pp. 915-919.

[14] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, pp. 82-93, Jan. 1990.

[15] T. R. Fischer, M. W. Marcellin, and M. Wang, "Trellis coded vector quantization," *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1551–1566, Nov. 1991.

[16] S. Kang, Y. Shin and T. R. Fischer, "Low-complexity predictive trellis coded quantization of speech line spectral frequencies," *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 2070–2079, Jul. 2004.

[17] J. Park and S. Kang, "Block constrained trellis coded vector quantization of LSF parameters for wideband speech coders," *ETRI Journal*, vol. 30, no. 5, pp. 738–740, Oct. 2008.

[18] A. Vasilache, A. Ramo, H. Sung, S. Kang, J. Kim, and E. Oh, "Flexible spectrum coding in the 3GPP EVS codec," *Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 2015, pp. 5878–5882.

[19] Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description, (Release 12), 3GPP TS 26.445, version 12.1.0, Dec. 2014.

[20] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[21] F. Lahouti and A. K. Khandani, "Quantization of LSF parameters using a trellis modeling," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 5, pp. 400–412, Sep. 2003.

[22] J. Skoglund and J. Linden, "Predictive VQ for noisy channel spectrum coding: AR or MA?," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 1997, pp. 1351–1354.

[23] J. S. Garofolo, L. Lamel, W. M. Fisher and D. S. Pallett, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," National Institute of Standards and Technology (NIST), Gaithersburgh, MD, 1988.

[24] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, May 1991, pp. 641–644.

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

[26] Y. Li and S. Kang, "Artificial bandwidth extension using deep neural network-based spectral envelope estimation and enhanced excitation estimation," *IET Signal Processing*, vol. 10, no. 4, pp. 422–427, Jun. 2016.

[27] Y. Li and S. Kang, "Deep neural network-based linear predictive parameter estimations for speech enhancement," *IET Signal Processing*, vol. 11, no. 4, pp. 469–476, Jun. 2017.

[28] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM–28, 1980, pp. 84–94.

[29] F. Lahouti, A. R. Fazel, A. H. Safavi-Naeini, and A. K. Khandani, "Single and double frame coding of speech LPC parameters using a lattice-based quantization scheme," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1624–1632, Sep. 2006.

[30] T. Eriksson, J. Linden, and J. Skoglund, "Exploiting inter-frame correlation in spectral quantization: a study of different memory VQ schemes," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, May 1996, pp. 765–768.