INTRA-PREDICTIVE SWITCHED SPLIT VECTOR QUANTIZATION OF SPEECH SPECTRA

Miguel Arjona Ramírez

University of São Paulo

ABSTRACT

Vector quantization (VQ) of speech spectral vectors has been improved by techniques such as split VQ (SVQ), vector transforms and direction switching. This paper proposes Intra-Predictive Switched SVQ (IPSSVQ) with direction switching by a Gaussian Mixture Model (GMM), using at the frame level the prediction-based lower-triangular transform (PLT), which has lower complexity than the Karhunen-Loève transform (KLT). It is shown that equivalent results to GMM KLT SSVQ may be obtained in the quantization of line spectral frequency (LSF) vectors from wideband speech signals, such as transparent coding throughout the range from 46 bit/frame to 41 bit/frame, with about three-fourths as much operational complexity.

Index Terms— vector quantization, intra-predictive quantization, prediction-based lower-triangular transform, Karhunen-Loève transform, line spectral frequencies.

1. INTRODUCTION

Vector quantization provides coding advantages over scalar quantization (SQ) that have been cleverly used for efficient implementations. The space-filling advantage (Table 1) increases with vector dimension at the expense of search complexity [1], which grows exponentially for full searches [2].

Table 1. Scalar and vector quantization properties and corresponding VQ advantages.

VQ property	SQ property	VQ advantage
linear dep(*)	linear dep(*)	memory
nonlinear dep(*)	none	memory and space-filling
pdf shape	pdf shape	shape
dimensionality	none	space-filling

(*) "dep" stands for "dependence".

A favorable trade-off of space-filling advantage for reduced complexity involves the use of product codebooks, whose component codes are concatenated to make up the complete VQ code. A very important subclass of product VQ for speech spectral parameters such as LSF vectors is split VQ, introduced by Paliwal and Atal [3].

Split VQ (SVQ) partitions the input vector into a number of subvectors of smaller dimension which are quantized by separate quantizers. It is noted that SQ is a limiting case of SVQ, where each subvector is a single vector component.

However, there is a split loss in SVQ because the dependencies among vector components in different splits go undetected [4]. Both SQ and SVQ improve performance if vector components are decorrelated prior to quantization since this removes the linear dependencies or correlations between them. In fact, an interesting scheme emerges by using SQ for quantizing the transformed components. Further, if uniform scalar quantizers are used after companding, a good trade-off is struck between performance and low complexity [5] with the additional possibility of scalable coding.

Another procedure that caters for the dependencies among components, but including their nonlinear part, is clustering of whole vectors prior to splitting. This initial quantization may be performed by full-dimension VQ, which leads to switched SVQ (SSVQ) [6, 7], or it may take the form of joint probability density function (pdf) modeling through a Gaussian mixture model (GMM) [5]. The latter has been shown to be a better solution, referred to as GMM-based SSVQ [8].

Nevertheless, in order to keep the complexity manageable, the number of clusters or switching directions in SSVQ is low, typically 8 or 16, which is insufficient to account for all dependencies among vector components. The correlation among the vector components lying in a given switching region is absorbed by their Karhunen-Loève transform (KLT) in the quantization that has been proposed as GMM-based KLT-domain SSVQ [9].

But the computation of the KLT involves eigenvalue decomposition in the training phase and its eigenvectors are not sparse, demanding more operations in the coding phase. Thus, a simpler decorrelating transform is desirable such as the prediction-based lower triangular transform (PLT) [10], which is proposed here for this application.

The fact that SVQ has been proven to improve quantizer performance, leading to GMM-KLT-SSVQ [9], and the ad-

The author acknowledges the support received from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant no. 307633/2011-0 and from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant no. 2012/24789-0. E-mail: miguel@lps.usp.br.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

vantages of VQ over SQ have motivated our research on SVQ for GMM Intra-Predictive quantization.

2. THE PREDICTION TRANSFORM

The prediction-based lower triangular transform diagonalizes the covariance matrix just as the KLT. However, its basis vectors are sparse, that is, the prediction transformed vector is

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} \tag{1}$$

for random source vector $\boldsymbol{x}$, where the direct PLT matrix is the unit-diagonal lower triangular matrix $\boldsymbol{B}$, resulting in the covariance matrix

$$\begin{aligned} \boldsymbol{R}_{yy} &= E\left[\boldsymbol{y}\boldsymbol{y}^T\right] \\ &= \boldsymbol{B}E\left[\boldsymbol{x}\boldsymbol{x}^T\right]\boldsymbol{B}^T \\ &= \boldsymbol{B}\boldsymbol{R}_{xx}\boldsymbol{B}^T, \end{aligned} \tag{2}$$

written as its Cholesky decompostion. The diagonal entries of $\boldsymbol{R}_{yy}$ are the backward prediction error variances $[\boldsymbol{R}_{yy}]_{ii} = \beta_i$ for $i = 0, 1, \ldots, p-1$, where p is the dimension of $\boldsymbol{x}$.

In [10] the PLT is introduced for sampled data vectors, whose covariance matrix is Toeplitz symmetric, whereas we are interested in LSF vectors with covariance matrix which is just symmetric but not Toeplitz. Therefore, the direct PLT matrix in Eq. (2) may be obtained by means of the covariance method of linear prediction (LP) [11], but not by the autocorrelation method as in the case of sampled data vectors.

The direct transform matrix for a p-dimensional source is the $p \times p$ unit-diagonal lower triangular matrix

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ b_{10} & 1 & \cdots & \cdots & 0 \\ b_{20} & b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p-1,0} & b_{p-1,1} & \cdots & \cdots & 1 \end{bmatrix} \tag{3}$$

where the nonzero entries in the mth row are the coefficients of the mth-order backward whitener

$$B_m(z) = z^{-m} + \sum_{i=0}^{m-1} b_{mi} z^{-i}, \tag{4}$$

for $m = 0, 1, \ldots, p-1$, where z^{-1} indicates the unit vector-component delay operator so that the no-delay component is x_0 and the $p-1$ unit delayed component is x_{p-1}. So, the inverse transform matrix

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ s_{10} & 1 & \cdots & \cdots & 0 \\ s_{20} & s_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{p-1,0} & s_{p-1,1} & \cdots & \cdots & 1 \end{bmatrix} \tag{5}$$

has the same structure, which enables its factorization into

$$\boldsymbol{S} = \boldsymbol{S}_1 \cdot \boldsymbol{S}_2 \cdot \ldots \cdot \boldsymbol{S}_{p-1}, \tag{6}$$

where $\boldsymbol{S}_m$ is the elementary matrix constructed from the $p \times p$ identity matrix $\boldsymbol{I}_p$ by replacing row m with the corresponding entries in $\boldsymbol{S}$ for $m = 1, 2, \ldots, p-1$. These elementary matrices may be inverted very efficiently by simply changing the algebraic signs of the entries in the lower triangle, while the unit values in the main diagonal remain untouched [10]. In this way, the direct transform matrix may be obtained factorized as

$$\boldsymbol{B} = \boldsymbol{S}_{p-1}^{-1} \cdot \boldsymbol{S}_{p-2}^{-1} \cdot \ldots \cdot \boldsymbol{S}_1^{-1}. \tag{7}$$

However, unlike the KLT, the PLT is not an orthogonal transform and the quantization error gain is greater than unity for direct filterbank implementations. Still, the factorizations (7) and (6) allow the splitting of the analysis and synthesis filterbanks into a ladder of elementary filterbanks. If, additionally, the typical scalar quantizer bank is split such that scalar quantizer $Q_m[\cdot]$ immediately follows elementary analysis bank $\boldsymbol{S}_m^{-1}$ for $m = 1, 2, \ldots, p-1$ and quantization starts by $\tilde{y}_0 = Q_0[x_0]$, then a minimum noise structure (MINLAB) is obtained as shown in Fig. 1. The analysis-quantize-synthesis structure just described was originally proposed as MINLAB(I) [10], which is more efficient than MINLAB(II) in this case.

It may be easily verified as follows that MINLAB(I) is a unit coding gain structure despite the fact that the PLT is nonunitary. Consider the quantization of the mth component

$$y_m = -s_{m0}\tilde{y}_0 - s_{m1}\tilde{y}_1 - \cdots - s_{m,m-1}\tilde{y}_{m-1} + x_m \tag{8}$$

as $\tilde{y}_m = Q_m[y_m]$ and its reconstruction as

$$\tilde{x}_m = s_{m0}\tilde{y}_0 + s_{m1}\tilde{y}_1 + \cdots + s_{m,m-1}\tilde{y}_{m-1} + \tilde{y}_m. \tag{9}$$

By adding Eqs. (8) and (9), we find that

$$y_m - \tilde{y}_m = x_m - \tilde{x}_m, \tag{10}$$

thus confirming that the reconstruction error equals the quantization error. This process may be streamlined if we first obtain the inverse transform matrix $\boldsymbol{S}$ by means of the Cholesky decompostion of $\boldsymbol{R}_{xx}$

$$\boldsymbol{R}_{xx} = \boldsymbol{S}\boldsymbol{R}yy\boldsymbol{S}^T \tag{11}$$

instead of decomposition (2).

3. INTRA-PREDICTION AND SCALAR QUANTIZATION

At first, the input vector space $\mathbb{R}^p$ is classified by means of the GMM of its joint pdf into N clusters $\mathbb{C}^{(n)}$ for $n = 1, 2, \ldots, N$, such that $\cup_{n=1}^{N}\mathbb{C}^{(n)} = \mathbb{R}^p$.

For scalar quantization, intra-prediction as applied in IPSSQ includes the PLT and the computation of the prediction residual subvector that results from the previous quantized components according to Eq. (8). This procedure is made possible due to the lower triangular structure of the prediction transform matrices.

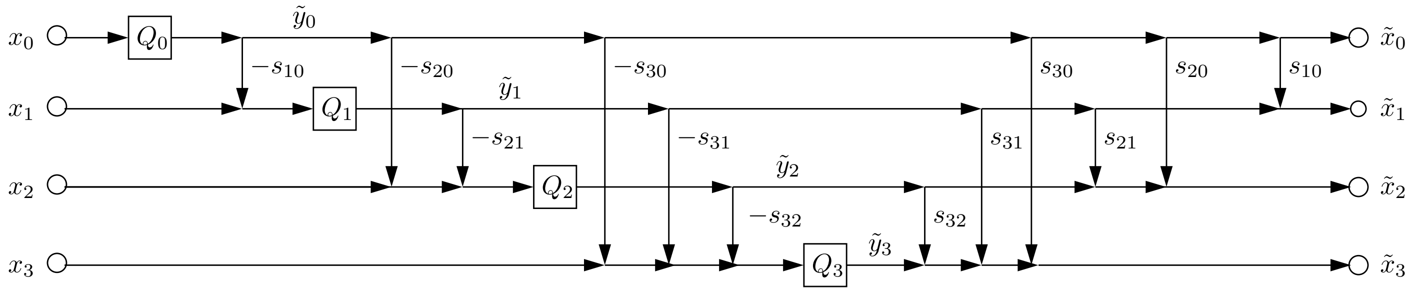

Fig. 1. Minimum noise structure I – MINLAB(I) – for intra-predictive quantization, including the inverse PLT at the right.

For a frame bit rate B, the remaining rate after preclassification is $B_q = B - \log_2 N$. Bit allocation is driven by the prediction error variances so that the optimal fixed bit rate for the ith component is

$$B_{qi} = \frac{B_q}{p} + \frac{1}{2}\log_2 \frac{\beta_i}{\left(\prod_{j=0}^{p-1}\beta_j\right)^{1/p}} \tag{12}$$

for $i = 0, 1, \ldots, p-1$, which are rounded to the nearest integer and adjusted so that $\sum_{i=0}^{p-1} B_{qi} = B_q$.

A weighted square distortion $d(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ is computed between the input vector $\boldsymbol{x}$ and its quantized version $\tilde{\boldsymbol{x}}$ as

$$d(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = (\boldsymbol{x} - \tilde{\boldsymbol{x}})^T \boldsymbol{D} (\boldsymbol{x} - \tilde{\boldsymbol{x}}) \tag{13}$$

where $\boldsymbol{D}$ is a diagonal matrix. In LSF vector quantization we use the sensitivity matrix of the log spectral distortion (SD) measure $d_{SD}(\cdot, \cdot)$ with entries

$$d_{ij} = \left.\frac{\partial^2 d_{SD}(\boldsymbol{x}, \tilde{\boldsymbol{x}})}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}=\tilde{\boldsymbol{x}}}, \tag{14}$$

which may be efficiently computed [12] from the predictor coefficients associated to LSF vector $\boldsymbol{x}$. Furthermore, it converges to a diagonal matrix in the high-rate quantization limit. That is why its diagonal is used in the diagonal matrix for the weighted measure, in a sense approximating the log SD.

4. INTRA-PREDICTION AND VECTOR QUANTIZATION

In SVQ the input vector is split into M subvectors such that subvector $\boldsymbol{x}_k$ lies in subspace $\mathbb{R}^{p_k}$ for $k = 1, 2, \ldots, M$ with $\sum_{k=1}^{M} p_k = p$ and $\mathbb{R}^p = \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \cdots \times \mathbb{R}^{p_M}$.

A further constraint of the PLT is the distribution of scalar quantizers in order to ensure unity quantization error gain as shown in Section 2. In order to comply with this condition, the codebooks are designed for SQ and composed by Cartesian products over each split span. For split k, the codebook is

$$\boldsymbol{\mathcal{C}}^{(k)} = \mathcal{C}_0^{(k)} \times \mathcal{C}_1^{(k)} \times \cdots \times \mathcal{C}_{p_k-1}^{(k)}. \tag{15}$$

However, it was found that the component codebooks have to be designed specifically for SVQ because the optimal bit allocation specified by Eq. (12) is not the best one for SVQ when added up over the components in the split. In fact, by imposing an upper bound on the split bit rate for efficiency, better results were found as reported in Section 5. Thus, a dimensionality advantage of SVQ over SQ (see Table 1) has been unveiled in connection with complexity reduction.

The extension of the analyze-quantize-transform Eq. (8) of SQ to the SVQ case, given the quantized lower subvectors $\tilde{\boldsymbol{u}}^{(k)} = \begin{bmatrix} \tilde{\boldsymbol{y}}^{(1)^T} & \tilde{\boldsymbol{y}}^{(2)^T} & \cdots & \tilde{\boldsymbol{y}}^{(k-1)^T} \end{bmatrix}^T$, involves the computation of the transformed subvector for split k as

$$\boldsymbol{y}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{S}^{(k)}\tilde{\boldsymbol{u}}^{(k)}, \tag{16}$$

where row i in $\boldsymbol{y}^{(k)}$,$\boldsymbol{x}^{(k)}$ and the $p_k \times \sum_{m=0}^{k-1} p_m$ matrix block $\boldsymbol{S}^{(k)}$ are extracted from row $\sum_{m=1}^{k-1} p_m + i$ in $\boldsymbol{y}$, $\boldsymbol{x}$ and the leftmost block of $\boldsymbol{S}$, respectively.

Next, the transformed subvector in Eq. (16) is vector quantized as

$$\tilde{\boldsymbol{y}}^{(k)} = \boldsymbol{Q}_k\left[\boldsymbol{y}^{(k)}\right], \tag{17}$$

where $\boldsymbol{Q}_k[\cdot]$ is the vector quantizer for split k. The subvector for split k is reconstructed as

$$\tilde{\boldsymbol{x}}^{(k)} = \tilde{\boldsymbol{y}}^{(k)} + \boldsymbol{S}^{(k)}\tilde{\boldsymbol{u}}^{(k)}. \tag{18}$$

Comparison of Eqs. (16) and (18) confirms that the SVQ structure proposed is a unit quantization error gain implementation.

5. CODING RESULTS

For the tests, LSF vectors were extracted from the wideband speech signals in the TIMIT database [13] by the adaptive multirate wideband (AMR-WB) speech coder [14] at a frame rate of 50 Hz and had their mean values subtracted. So 705,580 training vectors and 257,852 test vectors were obtained. Mean values were evaluated over the training database.

All tests have been performed with the weighted square distortion measure using the log SD sensitivity matrix in the high-rate approximation as described in Section 3. The frame bit rates under study range from 46 bit/fr down to 41 bit/fr.

The criteria for transparent coding established by [3] for narrowband speech and validated by [15] for wideband speech are mean SD around 1 dB, no outlying frame above 4 dB and less than 2% of outlying frames with SD in the range of 2–4 dB. Baseline results for SVQ are referred to [16], where a transparent coding threshold has been estimated at 46 bit/fr.

Transforms can remove some linear dependence before splitting but that is not a significant effect on its own. Their effect is more significant when switching is used first. So transform quantizers are introduced with the KLT after 8-direction GMM switching and followed by SQ in Table 2, where a great improvement in average performance is observed by the extension of coding transparency to the whole rate range. The equivalent GMM Intra-Predictive SQ results are shown in Table 3 with improved outlier performance.

Table 2. Performance of GMM KLT switched scalar quantization for 16-dimensional LSF vectors.

Bit rate	Mean	Outliers	
Per frame	SD	2 – 4 dB	> 4 dB
(bit/frame)	(dB)	(%)	(ppm)
46	0.824	0.36	12
45	0.855	0.39	12
44	0.882	0.55	12
43	0.916	0.72	12
42	0.950	0.91	16
41	0.981	1.14	19

Table 3. Performance of GMM Intra-Predictive switched scalar quantization for 16-dimensional LSF vectors.

Bit rate	Mean	Outliers	
Per frame	SD	2 – 4 dB	> 4 dB
(bit/frame)	(dB)	(%)	(ppm)
46	0.815	0.32	0
45	0.854	0.32	0
44	0.893	0.56	0
43	0.926	0.69	4
42	0.974	1.03	4
41	1.015	1.39	4

The improvement in performance obtained by using SVQ in this context can be seen in Table 4, where GMM KLT SSVQ [9] shows improved outlier and average performances. Also, the enhancement provided by IPSSVQ, constrained to a maximum single split rate of 8 bit/fr, can be observed in Table 5 with superior far outlier performance and improved average performance, even though it is below that of GMM KLT SSVQ, but still keeping transparency throughout the rate range under study.

Operational complexity for encoding GMM SSVQ without transforms is $6Mp + \sum_{i=1}^{M} \sum_{j=1}^{S} (4D_j - 1)\, 2^{B_{ij}}$ in flop/fr for M clusters, transform length p, S splits, dimension D_j for split j and bit rate per frame B_{ij} for split j in cluster i. For GMM IPSSVQ an additional analysis-synthesis complexity of $2Mp^2 - 2Mp$ applies and for GMM KLT SSVQ the corresponding additional complexity is $4Mp^2 - 2Mp$. Using the bit allocation for each case, the values for the complexities are shown in Table 6, where it is important to observe that the operational complexity for GMM IPSSVQ is around 3/4 that of GMM KLT SSVQ.

Table 4. Performance of GMM KLT SSVQ for 16-dimensional LSF vectors in (2,2,2,2,4,4)-dimensional splits.

Bit rate	Mean	Outliers	
Per frame	SD	2 – 4 dB	> 4 dB
(bit/frame)	(dB)	(%)	(ppm)
46	0.753	0.14	4
45	0.782	0.18	4
44	0.818	0.25	4
43	0.854	0.34	4
42	0.888	0.32	4
41	0.920	0.58	4

Table 5. Performance of GMM Intra-Predictive SSVQ for 16-dimensional LSF vectors in (2,3,3,3,3,2)-dimensional splits.

Bit rate	Mean	Outliers	
Per frame	SD	2 – 4 dB	> 4 dB
(bit/frame)	(dB)	(%)	(ppm)
46	0.804	0.21	0
45	0.835	0.26	0
44	0.861	0.34	0
43	0.906	0.36	0
42	0.942	0.61	0
41	0.963	0.65	4

6. CONCLUSION

A novel GMM Intra-Predictive SSVQ has been proposed. Intra-prediction is performed by the PLT, which has been conveniently factored for split VQ so that quantization error gain is unity. The advantages of SVQ over SQ have been unveiled by combining SQ codebooks designed under a maximum bit rate constraint at the split level and by using proper spectral sensitivity weighting on coding. A lower complexity of around 3/4 as much as that of GMM KLT SSVQ has been

Table 6. Operational complexity of GMM Intra-Predictive SSVQ for 16-dimensional LSF vectors in (2,3,3,3,3,2)-dimensional splits compared to that of GMM KLT SSVQ for a (2,2,2,2,4,4)-dimensional partition.

Bit rate	IPSSVQ	KLT SSVQ	Ratio
Per frame (bit/frame)	Complexity (kflop/frame)	Complexity (kflop/frame)	IP to KLT (%)
46	86	135	64
45	82	117	70
44	75	106	71
43	74	90	82
42	70	86	81
41	63	81	78

achieved while maintaining coding transparent for LSF vectors from wideband speech within the range from 46 bit/frame through 41 bit/frame.

7. REFERENCES

[1] Tom D. Lookabaugh and Robert M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1020–1033, Sept. 1989.

[2] John Makhoul, Salim Roucos, and Herbert Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, no. 11, pp. 1551–1588, Nov. 1985.

[3] Kuldip K. Paliwal and Bishnu S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 3–14, Jan. 1993.

[4] Fredrik Nordén and Thomas Eriksson, "On split quantization of LSF parameters," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, Montreal, Canada, 2004, vol. 1, pp. 157–160.

[5] Anand D. Subramaniam and Bhaskar D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 2, pp. 130–142, Mar. 2003.

[6] Stephen So and Kuldip K. Paliwal, "Efficient product code vector quantisation using the switched split vector quantiser," *Digital Signal Process.*, vol. 17, pp. 138–171, 2007.

[7] Stephen So and Kuldip K. Paliwal, "Switched split vector quantisation of line spectral frequencies for wideband speech coding," in *Proc. Eur. Conf. Speech Communication and Technology (INTERSPEECH 2005 - EUROSPEECH)*, Lisbon, Portugal, 2005, pp. 2705–2708.

[8] S. Chatterjee and T. V. Sreenivas, "Gaussian mixture model based switched split vector quantization of LSF parameters," in *Proc. IEEE Int. Symp. Signal Process. Inf. Tech.*, Cairo, Egypt, 2007, pp. 1054–1059.

[9] Yoonjoo Lee, Wonjin Jung, and Moo Young Kim, "GMM-based KLT-domain switched-split vector quantization for LSF coding," *IEEE Signal Processing Lett.*, vol. 18, pp. 415–418, July 2011.

[10] See-May Phoong and Yuan-Pei Lin, "Prediction-based lower triangular transform," *IEEE Trans. Signal Processing*, vol. 48, no. 7, pp. 1947–1955, July 2000.

[11] Bishnu S. Atal and Suzanne L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, no. 2, pp. 637–655, 1971.

[12] William R. Gardner and Bhaskar D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 5, pp. 367–381, Sept. 1995.

[13] John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, Nancy L. Dahlgren, and Victor Zue, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium, 1993, `http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1`.

[14] Bruno Bessette, Redwan Salami, Roch Lefebvre, Milan Jelínek, Jani Rotola-Pukkila, Janne Vainio, Hannu Mikkola, and Kari Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov. 2002.

[15] G. Biundo, S. Pauletti, M. Ansorge, F. Pellandini, and P.A. Farine, "Design techniques for spectral quantization in wideband speech coding," in *Proc. 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, Budapest, Hungary, Oct. 2002, vol. 1, pp. 114–119.

[16] M. Arjona Ramírez, "Vector quantization with renormalized splits for wideband speech," in *Proc. of DSP 2011 17th Int. Conf. on Digital Signal Processing*, Corfu, Greece, 2011, pp. 1–4.