

Optimized Subvector Processing in Split Vector Quantization

Miguel Arjona Ramírez

Abstract—Split vector quantization (SVQ) is efficient but suboptimal. Here a renormalization process is proposed for intraframe splitting and joining of subvectors, which integrates gracefully with trained interframe prediction. Renormalization increases the availability of codevectors for the quantization of each subvector in ordered vectors such as the line spectral frequency (LSF) vectors. For 16-dimensional LSF vectors from wideband speech, renormalized SVQ (RSVQ) is shown to achieve a savings of 4 bit/frame over standard SVQ, reaching transparent coding at 42 bit/frame. Further, predictive RSVQ saves an additional 4 bit/frame for transparent coding down to 38 bit/frame.

Index Terms—split vector quantization, line spectral frequencies, vector prediction, speech analysis, speech coding.

I. INTRODUCTION

SPLIT vector quantization overcomes the curse of dimensionality inherent in vector quantization by splitting the vector into lower-dimensional subvectors. It is particularly efficient when the distortion measure used for quantization is separable. In particular, it is widely used for the quantization of line spectral frequency (LSF) vectors that represent the short-term spectral envelope of speech signals [1].

However, an amount of suboptimality remains that is referred to as the split loss [2]. It may be partially counteracted by classified vector quantization (VQ) [3] and by combining split VQ with multistage VQ [4].

In a different approach, a method is proposed that increases the availability of usable codevectors in the split codebooks when the vectors are ordered. An important case are the LSF vectors. In particular, wideband speech is used here to extract them.

The method consists of two actions. First, in the training or in the encoding phase, the bandwidth for the current split is normalized to cover the range defined by the previously coded LSFs in the neighboring splits. Second, the band spanned by the codevectors in the split is renormalized to close the gap between the left and right neighboring splits in the vector to be decoded. Previously, for narrowband speech, normalization of LSF vectors has been used per se [5] or combined with classified VQ [6].

Still, the performance may be improved by predictive RSVQ (renormalized SVQ), which introduces interframe coding along with the inherent intraframe coding in RSVQ. This

This work is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant no. 309249/2008-2 and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant no. 2009/18242-5. The author is with Dept. of Electronic Systems Engineering, Escola Politécnica, University of São Paulo, SP, Brazil, e-mail: miguel@lps.usp.br

further enhances the decrease in split loss provided by renormalization as explained in Section IV, which is followed by the results in training and performance evaluation in Section V right before the conclusion. Before that, RSVQ is described in Section III following the discussion about distortion measures in Section II.

II. DISTORTION MEASURES AND VECTOR PARTITIONING

The log spectral distortion (SD) is the most usual measure of distortion for assessing the quantization performance for parameter vectors that represent the spectral envelope of speech signals. This is due to its correlation with human perception, at least within the distortion range covered by usual rate range, and to the transparent coding rules set forth by Paliwal and Atal in their introduction of SVQ [1] for narrowband speech, which essentially state for transparent quantization that

- 1) The mean SD is about 1 dB.
- 2) There is no outlier exceeding 4 dB in SD.
- 3) The number of outliers having SD in the range from 2 dB to 4 dB is less than 2%.

Later, these criteria were verified to hold for wideband speech as well [3].

For LSF vectors \mathbf{f} and $\tilde{\mathbf{f}}$ with synthesis filters having power spectral densities (PSDs) $P(f)$ and $\tilde{P}(f)$, respectively, the SD is defined by

$$D_S(\mathbf{f}, \tilde{\mathbf{f}}) = \sqrt{\int_0^1 (10 \log_{10} P(f) - 10 \log_{10} \tilde{P}(f))^2 df}, \quad (1)$$

where f is the cycle frequency in cycle/sample.

However, Eq. (1) is too complex to be used for designing the codebook in the training phase of the quantizer and even on encoding and decoding because the full-dimension LSF vector is needed for computing its PDF. Besides, particularly for SVQ, a separable distortion measure contributes to an initial decrease in split loss. In general, a weighted Euclidean distance is used instead. In particular, we use the following dynamically weighted square measure [1]

$$d(\mathbf{f}, \tilde{\mathbf{f}}) = (\mathbf{f} - \tilde{\mathbf{f}})^T \mathbf{W} (\mathbf{f} - \tilde{\mathbf{f}}) \quad (2)$$

where \mathbf{W} is the diagonal weighting matrix with main diagonal entries

$$w_{ii} = P^r(f_i) \quad (3)$$

for $i = 1, 2, \dots, p$, where p is the dimension of the LSF vectors and $P^r(f_i)$ stands for the r th power of the PSD at frequency f_i . The exponent of the PSD is set to $r = 0.3$.

For split VQ, LSF vector $\mathbf{f} = \{f_i\}_{i=1}^p$ is partitioned as

$$\mathbf{f} = [\varphi_1^T \quad \varphi_2^T \quad \cdots \quad \varphi_\varsigma^T]^T,$$

where ς is the number of partitions or splits and the i th subvector consists of

$$\varphi_i = [f_{\delta_i} \quad f_{\delta_i+1} \quad \cdots \quad f_{\delta_i+D_i-1}]^T \quad (4)$$

with initial LSF index δ_i and dimension D_i . Overall, the beginning and end boundaries are $\delta_1 = 1$ and $\delta_\varsigma + D_\varsigma - 1 = p$. Likewise, the reconstructed LSF vector $\tilde{\mathbf{f}} = \{\tilde{f}_i\}_{i=1}^p$ is partitioned as

$$\tilde{\mathbf{f}} = [\tilde{\varphi}_1^T \quad \tilde{\varphi}_2^T \quad \cdots \quad \tilde{\varphi}_\varsigma^T]^T,$$

where the reconstructed subvector for the i th split after split vector quantizing with the i th codebook is

$$\tilde{\varphi}_i = [\tilde{f}_{\delta_i} \quad \tilde{f}_{\delta_i+1} \quad \cdots \quad \tilde{f}_{\delta_i+D_i-1}]^T \quad (5)$$

Now it is observed that the distortion measure in Eq. (2) is separable in split components, that is,

$$d(\mathbf{f}, \tilde{\mathbf{f}}) = \sum_{i=1}^{\varsigma} d_i(\varphi_i, \tilde{\varphi}_i), \quad (6)$$

where the distortion component due to the i th subvector is

$$d_i(\varphi_i, \tilde{\varphi}_i) = \sum_{j=\delta_i}^{\delta_i+D_i-1} w_{jj} (f_j - \tilde{f}_j)^2. \quad (7)$$

III. RENORMALIZATION OF SUBVECTORS

Renormalization enables a major reduction in split loss. As shown in Section II, by using a separable distortion measure, some amount of split loss is prevented from the outset. This means that there is no information about split loss in the cumulative split distortion that we can use in order to reduce the split loss any further.

Nonetheless, as we can see from the distribution of LSF vectors in the training database shown in Fig. 5, presented in Section V, the bands spanned by nearby splits overlap considerably, sometimes reaching the second neighbor.

Based on the observation above, it is postulated that much of the split loss originates when the effective size of the codebook is reduced by the enforcement of the stability relations

$$0 < \tilde{f}_1 < \tilde{f}_2 < \cdots < \tilde{f}_{p-1} < \tilde{f}_p < 1/2 \quad (8)$$

which exclude from the search the split codevectors that do not obey them. For instance, when coding the i th subvector, only those codevectors are considered for which it holds that

$$f_{\delta_i} > \tilde{f}_{\delta_i-1}, \quad (9)$$

that is, the lowest LSF in the split must be greater than the highest quantized LSF in the previous split.

One possibility for normalization of the i th split is to set its lower band edge to the quantized value of the largest LSF coefficient in the lower split of the current frame m in the training database, or of the frame to be encoded, as

$$f_{iL}(m) = \tilde{f}_{\delta_i-1}(m), \quad (10)$$

and let $f_{iU} = 1/2$ be the upper band edge. We will call this version of RSVQ the sequential RSVQ and it is represented in Fig. 1, which indicates that the original lower band edge is mapped to the normalized zero frequency and the Nyquist frequency is mapped to the normalized unity frequency. Next, keeping the relative positions within the split band, the LSF subvector to be quantized in the i th split is normalized as

$$f'_j(m) = \frac{f_j(m) - f_{iL}(m)}{f_{iU} - f_{iL}(m)} \quad (11)$$

for $j = \delta_i, \delta_i + 1, \dots, \delta_i + D_i - 1$. Then, the normalized subvector $\varphi'_i(m)$ is quantized as $\tilde{\varphi}'_i(m)$, selected from the codebook for the i th split.

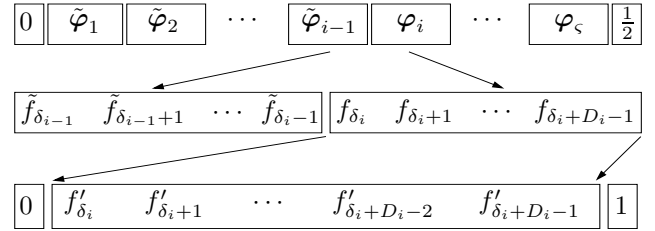


Fig. 1. Normalization as performed by sequential RSVQ at the i th split.

In the decoding phase, as shown in Fig. 2, the i th subvector is reinserted by mapping the normalized zero frequency to the lower band edge set by the highest frequency in the next lower split and by mapping the normalized unity frequency to the Nyquist frequency. More specifically, the lower band edge for renormalization of the LSF subvector $\varphi_i(m)$ is identical to the highest quantized LSF in the preceding split of the same LSF vector and is set as

$$\tilde{f}_{iL}(m) = \tilde{f}_{\delta_{i-1}}(m) \quad (12)$$

to be used for the renormalization that generates the quantized subvector $\tilde{\varphi}_i(m)$ as

$$\tilde{f}_j(m) = (f_{iU} - \tilde{f}_{iL}(m)) \tilde{f}'_j(m) + \tilde{f}_{iL}(m) \quad (13)$$

for $j = \delta_i, \delta_i + 1, \dots, \delta_i + D_i - 1$.

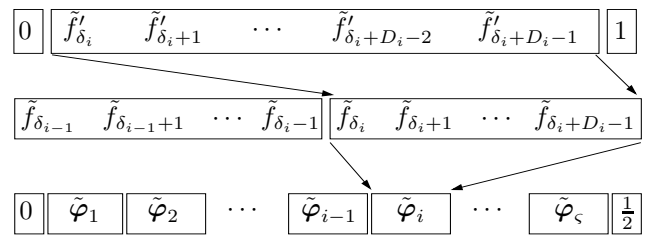


Fig. 2. Renormalization as performed by sequential RSVQ at the i th split.

Another sequencing order for normalization is used in interleaved RSVQ, which is anchored in the two splits neighboring the one under quantization as shown in Fig. 3, mapping the original lower band edge to the zero frequency and the original upper band edge to the unity frequency. More specifically, the lower band edge for the i th split is determined by Eq. (10), similarly to the sequential case, and the upper band edge is set as

$$f_{iU}(m) = \tilde{f}_{\delta_{i+1}}(m), \quad (14)$$

that is, the upper edge is the lowest LSF value in the following split.

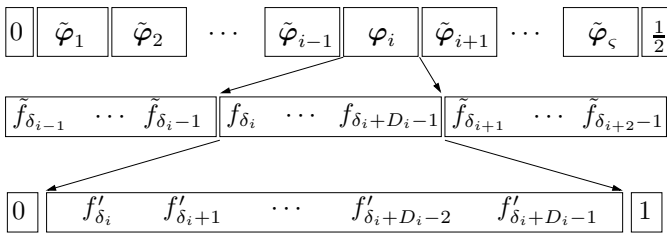


Fig. 3. Normalization as performed by interlaced RSVQ at the i th split.

Now, renormalizations in the decoding phase use the same assignment for the lower band edge of the i th split in the m th frame as described by Eq. (12), but the upper band edge now is adaptive and set by

$$\tilde{f}_{iU}(m) = \tilde{f}_{\delta_{i+1}}(m) \quad (15)$$

to be used in the renormalization of the selected codevector $\tilde{\varphi}'_i(m)$ according to

$$\tilde{f}'_j(m) = (f_{iU}(m) - f_{iL}(m)) \tilde{f}'_j(m) + f_{iL}(m) \quad (16)$$

for $j = \delta_i, \delta_i + 1, \dots, \delta_i + D_i - 1$. Naturally, both the $(i-1)$ th and the $(i+1)$ th split must have been quantized first with standard SVQ.

In an experiment reported in Section V, the sequential version is found to be superior to the interlaced one by a small margin.

IV. PREDICTIVE SPLIT QUANTIZATION

Nearby frames share a considerable amount of correlation that may be removed by a linear predictor. In fact, a linear predictor can remove covariance if the mean subvector is subtracted. We use the mean subvector $\bar{\varphi}_i$ for the training database to get the centered subvectors $\nu_i(m) = \varphi(m) - \bar{\varphi}$ for splits $i = 1, 2, \dots, \zeta$.

Now, for the centered subvectors, autocorrelation coefficients are enough for prediction. A first-order vector-valued moving-average predictor is used around the subvector quantizer as shown in Fig. 4, where α_i is the scalar prediction coefficient, $\tilde{\nu}_i(m)$ is the reconstructed centered subvector for the m th frame, $r_i(m)$ is the prediction residual subvector and $\tilde{r}_i(m)$ is the codevector selected by the split vector quantizer. Finally, the reconstructed LSF subvector is found by restoring the mean subvector as

$$\tilde{\varphi}_i(m) = \tilde{\nu}_i(m) + \bar{\varphi}_i. \quad (17)$$

In predictive RSVQ, the vector processing in the quantizer-predictor loop works in much the same way as for standard SVQ except for the subvectors to be quantized, which are the normalized subvectors $\varphi'_i(m)$ instead of the original subvectors $\varphi_i(m)$.

Given the structure drawn in Fig. 4, starting from the corresponding memoryless quantizer, the prediction coefficient

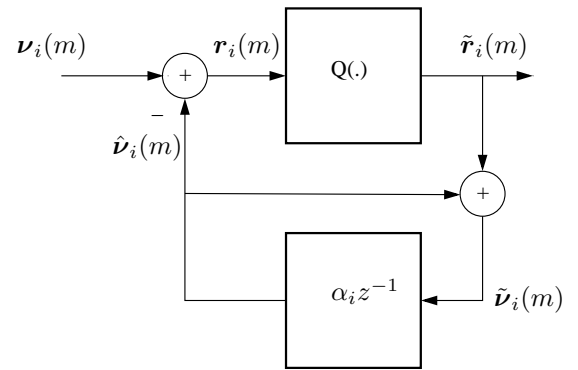


Fig. 4. First-order predictive vector quantizer for centered subvectors in the i th split.

α_i is adjusted iteratively by minimizing the norm square of the residual vector

$$r_i(m) = \nu_i(m) - \alpha \tilde{\nu}_i(m-1) \quad (18)$$

over the N frames in the training database. This leads to the prediction coefficient estimate

$$\alpha_i = \frac{R_{\tilde{\nu}_i \nu_i}(1)}{R_{\tilde{\nu}_i \tilde{\nu}_i}(0)} \quad (19)$$

by means of the correlation coefficients, which are computed as

$$R_{\tilde{\nu}_i \nu_i}(1) = \sum_{j=1}^{D_i} \sum_{m=0}^{N-2} \tilde{\nu}_{ij}(m) \nu_{ij}(m+1) \quad (20)$$

$$R_{\tilde{\nu}_i \tilde{\nu}_i}(0) = \sum_{j=1}^{D_i} \sum_{m=0}^{N-2} \tilde{\nu}_{ij}^2(m), \quad (21)$$

where N is the number of frames in the training database, D_i is the i th split dimension and the centered subvector in the i th split is represented in terms of its entries as

$$\tilde{\nu}_i = [\tilde{\nu}_{i1} \quad \tilde{\nu}_{i2} \quad \dots \quad \tilde{\nu}_{i,D_i}]^T. \quad (22)$$

In training sessions reported in Section V, the split prediction coefficients have converged to values lying between 0.8 and 0.9.

V. QUANTIZER TRAINING AND EVALUATION

The designs and tests reported here used the TIMIT speech database[7]. The speech signals were segmented with asymmetric Hamming windows at a rate of 50 Hz using the 3GPP AMR wideband coder [8] to provide the linear prediction coefficients used to obtain the LSF vectors. Then the training partition, having 705,580 frames, was used for design and the test partition with 257,852 frames was assigned to the quantization tests.

A preliminary test was performed to find out which of the two sequences of split quantization performs better. The test was performed at 46 bit/frame with (9,10,9,9) bit/split for the 5 splits having dimensions (3,3,3,3,4), respectively, with results displayed in Table I. Both RSVQ versions perform considerably better than SVQ but the sequential RSVQ version

TABLE I

COMPARISON AMONG SEQUENTIAL AND INTERLACED RSVQ AND STANDARD SVQ FOR 16-DIMENSIONAL LSF VECTORS IN (3,3,3,3,4)-DIMENSIONAL SPLITS AT A RATE OF 46 BIT/FRAME, INCLUDING MEAN LOG SPECTRAL DISTORTION AND TWO CLASSES OF OUTLIERS.

Algorithm	Mean SD (dB)	Outliers	
		2 - 4 dB (%)	> 4 dB (ppm)
Standard SVQ	0.924	0.46	0
Sequential RSVQ	0.821	0.22	0
Interlaced RSVQ	0.837	0.22	0

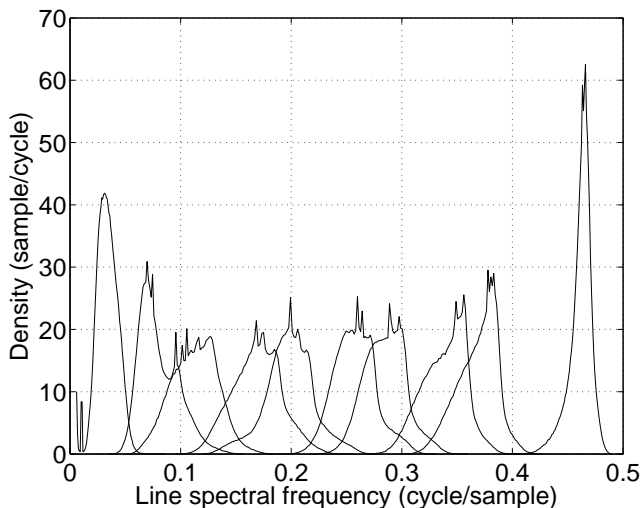


Fig. 5. Marginal density functions for endpoints in splits 1 through 5 over the training partition of the database.

is found to be slightly better than interlaced RSVQ. So it was chosen for the following tests.

The distributions in the training database of the LSF values at the endpoints of the splits are shown superimposed in Fig. 5, where the overlapping is seen to be significant.

The tests involving memoryless quantizers covered the range of rates from 40 bit/frame to 46 bit/frame using the same dimensional splitting above and a summary of their results is shown in Tables II and III. It is noted that SVQ performs transparently from 46 bit/frame upwards since the number of outliers > 4 dB is nonzero for the rate just below. On the other hand, for RSVQ transparent coding extends down to 42 bit/frame since the mean SD rises above 1 dB at 41 bit/frame. Therefore, RSVQ saves 4 bit/frame over SVQ for memoryless quantizers.

For predictive quantizers, the results are presented in Tables IV and V, where a decrease of around 0.3 dB in mean SD can be observed at 46 bit/frame due to the predictor.

Transparent coding starts at 41 bit/frame for predictive SVQ and at 38 bit/frame for predictive RSVQ since outliers above 4 dB make their appearance when the rate is decreased by 1 bit/frame in both cases. Therefore, prediction saves 5 bit/frame for SVQ and makes the savings for RSVQ reach 8 bit/frame. It is further observed that the 38 bit/frame

TABLE II

PERFORMANCE OF STANDARD SPLIT VECTOR QUANTIZATION FOR 16-DIMENSIONAL LSF VECTORS IN (3,3,3,3,4)-DIMENSIONAL SPLITS, INCLUDING MEAN LOG SPECTRAL DISTORTION AND TWO CLASSES OF OUTLIERS.

Bit rate		Mean SD (dB)	Outliers	
Per frame (bit/frame)	Per split (bit/split)		2 - 4 dB (%)	> 4 dB (ppm)
40	(8,8,8,8,8)	1.175	2.55	19
41	(8,9,8,8,8)	1.127	1.92	19
42	(8,9,9,8,8)	1.072	1.15	40
43	(8,9,9,9,8)	1.046	1.09	40
44	(9,9,9,9,8)	1.019	1.01	40
45	(9,9,9,9,9)	0.962	0.59	40
46	(9,10,9,9,9)	0.924	0.46	0

TABLE III

PERFORMANCE OF RENORMALIZED SPLIT VECTOR QUANTIZATION FOR 16-DIMENSIONAL LSF VECTORS IN (3,3,3,3,4)-DIMENSIONAL SPLITS, INCLUDING MEAN LOG SPECTRAL DISTORTION AND TWO CLASSES OF OUTLIERS.

Bit rate		Mean SD (dB)	Outliers	
Per frame (bit/frame)	Per split (bit/split)		2 - 4 dB (%)	> 4 dB (ppm)
40	(8,8,8,8,8)	1.047	1.29	0
41	(8,9,8,8,8)	1.003	0.96	0
42	(8,9,9,8,8)	0.968	0.75	0
43	(8,9,9,9,8)	0.938	0.59	0
44	(9,9,9,9,8)	0.907	0.51	0
45	(9,9,9,9,9)	0.855	0.26	0
46	(9,10,9,9,9)	0.821	0.22	0

transparent coding threshold for predictive RSVQ lies just 3 bit/frame above the informal lower bound derived in [3].

VI. CONCLUSION

Split vector quantization is the most efficient suboptimal method when used in isolation, particularly for LSF memoryless vector quantization. A renormalization process has been proposed for wideband LSF SVQ which reduces its split loss by an amount of 4 bit/frame as measured by the transparent coding threshold rate. Further, one-step predictive VQ has been used and proven capable of reducing the transparent coding

TABLE IV

PERFORMANCE OF PREDICTIVE STANDARD SVQ FOR 16-DIMENSIONAL LSF VECTORS IN (3,3,3,3,4)-DIMENSIONAL SPLITS, INCLUDING MEAN LOG SPECTRAL DISTORTION AND TWO CLASSES OF OUTLIERS.

Bit rate		Mean SD (dB)	Outliers	
Per frame (bit/frame)	Per split (bit/split)		2 - 4 dB (ppm)	> 4 dB (ppm)
39	(8,8,8,8,7)	0.829	3021	23
40	(8,8,8,8,8)	0.778	1703	16
41	(8,9,8,8,8)	0.745	1288	0
42	(8,9,9,8,8)	0.723	981	0
43	(8,9,9,9,8)	0.702	803	0
46	(9,10,9,9,9)	0.616	244	0

TABLE V

PERFORMANCE OF PREDICTIVE RSVQ FOR 16-DIMENSIONAL LSF VECTORS IN (3,3,3,3,4)-DIMENSIONAL SPLITS, INCLUDING MEAN LOG SPECTRAL DISTORTION AND TWO CLASSES OF OUTLIERS.

Bit rate		Mean SD (dB)	Outliers	
Per frame (bit/frame)	Per split (bit/split)		2 – 4 dB (ppm)	> 4 dB (ppm)
37	(7,8,8,7,7)	0.796	2145	4
38	(7,8,8,8,7)	0.768	1869	0
39	(8,8,8,8,7)	0.740	1555	0
40	(8,8,8,8,8)	0.697	756	0
46	(9,10,9,9,9)	0.544	101	0

threshold by an additional 4 bit/frame, reaching transparent coding at 38 bit/frame.

REFERENCES

[1] K. K. Paliwal and B. 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.

[2] S. Chatterjee and T. V. Sreenivas, "Sequential split vector quantization of LSF parameters using conditional PDF," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, Honolulu, 2007, pp. IV–1101–IV–1104.

[3] S. So and K. K. Paliwal, "A comparative study of LPC parameter representations and quantisation schemes for wideband speech coding," *Digital Signal Processing*, vol. 17, pp. 114–137, 2007.

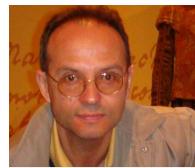
[4] B. Bessette, R. Salami, R. Lefebvre, M. Jelínek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov. 2002.

[5] A. M. . Smith, J. P. Ashley, M. A. Jusiuk, and W. Peng, "Normalization and polygon error detection for split VQ of line spectral frequencies," in *Proc. of IEEE Workshop on Speech Coding*, Delavan, 2000, pp. 123–125.

[6] H. Sung, C. Son, K. Lee, and E. Oh, "An efficient LSF quantization using dynamic bit allocation," in *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, Taipei, Taiwan, 2009, pp. 4141–4144.

[7] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium, 1993, <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1>.

[8] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; ANSI-C code," 3GPP TS 26.204 V7.1.0 (2007-09) Release 7, 2007.



Miguel Arjona Ramirez is Associate Professor at Escola Politécnica, University of São Paulo, where he is a member of the Signal Processing Laboratory. He received the E.E. degree from Instituto Tecnológico de Aeronáutica, Brazil, and the M.S., the Dr. and the Habilitation (Livre-Docência) degrees in Electrical Engineering from University of São Paulo, Brazil, in 1992, 1997 and 2006, respectively, and the Electronic Design Eng. degree from Philips International Institute, The Netherlands, in 1981. In 2008 he carried post-doctoral research in time-frequency speech analysis and coding at the Royal Institute of Technology in Stockholm, Sweden. He was Engineering Development Group Leader for Interactive Voice Response Systems (IVRs) for Itautec Informática, Brazil, where he served from 1982 to 1990. He is a Senior Member of the IEEE since 2000, a Member of the Brazilian Telecommunications Society (SBRt) and a Member of the IEICE. His research interests include signal compression, speech analysis, coding and recognition, and audio analysis and coding.