Event Location by Triangular Interpolation for Temporal Decomposition of Speech

Miguel Arjona Ramírez

***Abstract*— For low-bit-rate coding and synthesis the evolution of spectral parameters is a source of redundancy to be considered. A triangular interpolation spectral measure (TRISM) is proposed as the basis for an open-loop event location criterion for low-delay temporal decomposition (TD). TRISM comes as an improvement in linear interpolation error measurement over the spectral transition measure (STM). While STM is heuristic and presupposes asymmetric event functions, TRISM is a minimum square interpolation error based on symmetric functions. Minimun TRISM (MINTRISM) TD interpolates up to 13 frames between adjacent events at a mean event rate of 15 Hz and interpolation error level equivalent to that of standard low-bit-rate speech coders. The MINTRISM criterion is also a more stable solution to the location of events and determination of their number than previous global and local TD methods.**

***Index Terms*— Linear predictive coding, speech coding, temporal decomposition, interpolation, autoregressive processes.**

I. Introduction

THE representation of speech spectral features plays a central role in speech coding, synthesis and recognition. Each spectral vector represents the envelope of the average speech spectrum along a frame, which is a quasistationary segment of speech that lasts typically for some tens of milliseconds. The line spectral frequency (LSF) coefficients [1], [2], [3] are the representation of choice for the spectral vectors in speech coding since they are very robust parameters against quantization and interpolation errors. For a pth-order linear prediction (LP) analysis, the LSFs constitute the complete set of p resonant frequencies of the lossless vocal tract model under both alternative conditions of open and closed termination at the glottis. The LSF values range over the doubly-open interval $(0, \pi)$ radians per sample, that is, from DC to the Nyquist frequency.

Variable-rate interpolation of target spectral vectors is implemented in various methods known as temporal decomposition [4]. The temporal decomposition of parameter tracks involves the location of event centers in the analysis phase when event targets are sampled at event center locations and then refined. The set of frames that lie between an event center inclusively and the next one exclusively is called a superframe. In the synthesis or recognition phase, event targets are interpolated by means of event functions in order to reconstruct the parameter tracks.

Manuscript received November 30, 2009; revised May 3, 2010. 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 Grants no. 2009/18242-5 and 2007/08288-2.
The author is with Dept. of Electronic Systems Engineering, Escola Politécnica, University of São Paulo, SP, Brazil, e-mail: miguel@lps.usp.br

Unconstrained TD incurs in long delays. While such algorithms are useful for speech recognition [5], store-and-forward messaging applications [6] and for compressing speech synthesis corpora [7], for two-way coding applications low-delay TD algorithms are necessary. That is why the TD algorithm to be described locates event centers one at a time and constrains event functions to a finite support that spans two consecutive intertarget intervals.

In particular, a spectral measure for event location in TD is proposed. The triangular interpolation spectral measure (TRISM) is based on interpolation error minimization and local slope minimization under the condition of a triangular event function. It is compared to the spectral feature transition rate (SFTR) [8], which reduces to the spectral transition measure (STM) [9] when the location window length is fixed.

There are no well established guidelines for acceptable interpolation distortion as there are for LSF quantization distortion. As a matter of fact, speech decoders usually interpolate frame rate vectors for a subframe resolution of one-fourth frame length, without accounting for the interpolation error incurred. An exceptional work in this respect was done by Paliwal [10], which may be taken as a baseline reference for the performance of uniform linear interpolation.

Besides, for low-bit-rate speech coding, weighted distortion measures across frame time and frequency[11], [12] should be considered.

II. Spectral measures for event location

The LSF evolution matrix $\mathbf{Y}$ contains, for each frame in the range $n = 0, 1, \ldots N-1$, p LSFs as column vector $\mathbf{y}(n)$. It is temporally decomposed, generating target matrix $\mathbf{A}$ and event matrix $\mathbf{\Phi}$, which may be used to estimate $\mathbf{Y}$ as

$$\hat{\mathbf{Y}} = \mathbf{A}\mathbf{\Phi}. \tag{1}$$

The columns $\mathbf{a}_j$ in matrix $\mathbf{A}$ for $j = 0, 1, \ldots, J-1$ are the target vectors, where J is the number of events. Event functions $\phi_j(n)$ for $n = 0, 1, \ldots, N-1$ and $j = 0, 1, \ldots, J-1$ are represented as vectors $\boldsymbol{\phi}_j$ whose transposes are the rows in matrix $\mathbf{\Phi}$.

In triangular TD event functions for error measurement are assumed to be linear interpolation functions which are symmetric around their center locations while STM-based TD uses asymmetric linear interpolation functions implicitly in STM evaluation.

In particular, local TD is performed between the current and the next event locations in a two-stage sequential process. The first stage involves the determination of the next event

location. Then, in the second stage, the next event target vector is determined along with the functions for the current and the next events in an iterative fashion.

In the first stage, event function $\phi_j(n)$ is assumed to reach its peak unity value at event center $C(j) = n_0$ and to be triangular and symmetric about it so that

$$\phi_j(n) = \alpha\,(n - n_0) + 1 \tag{2}$$

for $n_0 \le n \le n_0 + M$ and

$$\phi_j(n) = -\alpha\,(n - n_0) + 1 \tag{3}$$

for $n_0 - M \le n \le n_0$, where α is the attack slope. Along the kth LSF track, the interpolation error is $e_k(n) = y_k(n) - a_{kj}\phi_j(n)$ for $n_0 - M \le n \le n_0 + M$, where a_{kj} is the target LSF value. The square interpolation error along the kth LSF track for the window $n_0 - M \le n \le n_0 + M$ is

$$\begin{aligned} \varepsilon_{kj} &= \sum_{m=0}^{M} \{y_k\,(n_0 + m) - a_{kj}\,[\alpha m + 1]\}^2 \\ &+ \sum_{m=1}^{M} \{y_k\,(n_0 - m) - a_{kj}\,[\alpha m + 1]\}^2 . \end{aligned} \tag{4}$$

The joint interpolation error of all LSF tracks is $\varepsilon_j = \sum_{k=1}^{p} \varepsilon_{kj}$. Expanding Eq. (4), rearranging the result and casting it in vector notation, yields

$$\begin{aligned} \varepsilon_j &= \sum_{m=-M}^{M} \|\mathbf{y}\,(n_0 + m)\|^2 \\ &- 2\alpha \mathbf{a}_j^T \sum_{m=1}^{M} m\,[\mathbf{y}\,(n_0 + m) + \mathbf{y}\,(n_0 - m)] \\ &- 2\mathbf{a}_j^T \mathbf{y}\,(n_0) - 2\mathbf{a}_j^T \sum_{m=1}^{M} [\mathbf{y}\,(n_0 + m) + \mathbf{y}\,(n_0 + m)] \\ &+ 2\alpha^2 \|\mathbf{a}_j\|^2 \sum_{m=1}^{M} m^2 \\ &+ 2M(M+1)\alpha \|\mathbf{a}_j\|^2 + (2M+1)\|\mathbf{a}_j\|^2 . \end{aligned} \tag{5}$$

Imposing $\frac{\partial \varepsilon_j}{\partial \alpha} = 0$ for minimum square interpolation error, the slope of the event function turns out to be

$$\hat{\alpha} = \frac{\mathbf{a}_j^T \sum_{m=1}^{M} m\,[\mathbf{y}\,(n_0 + m) + \mathbf{y}\,(n_0 - m)]}{\|\mathbf{a}_j\|^2 \sum_{m=-M}^{M} m^2} - \frac{M(M+1)}{\sum_{m=-M}^{M} m^2}. \tag{6}$$

It is noticeable that $\alpha = \hat{\alpha}$ is really a unique minimum for $\varepsilon_j\,(\alpha)$ at given window length $2M + 1$ and location $n = n_0$ since, by differentiating twice (5) with respect to α, we get

$$\frac{\partial^2 \varepsilon_j}{\partial \alpha^2} = 4\|\mathbf{a}_j\|^2 \sum_{m=1}^{M} m^2, \tag{7}$$

which is strictly positive.

For event center location, the target vector $\mathbf{a}_j$ is identified to the central LSF vector $\mathbf{y}\,(n_0)$. The criterion proposed for determining event centers is the minimization over frame time of the triangularly fit slope $\hat{\alpha}(n)$. For a given location window length $2M+1$, this is equivalent to the determination of frame locations n that locally minimize the triangular interpolation spectral measure (TRISM), defined by

$$\begin{aligned} T_M\,(n) &= \left| \frac{\mathbf{y}^T(n) \sum_{m=1}^{M} m\,[\mathbf{y}\,(n+m) + \mathbf{y}\,(n-m)]}{\|\mathbf{y}(n)\|^2} \right. \\ &- \left. M(M+1) \right| , \end{aligned} \tag{8}$$

which consists of the scaled version of the absolute value of slope estimate (6).

In a previous local TD method, event functions are assumed to be linear and minimum event function slope is taken to be the manifestation of spectral stability, whose location is declared event center [8]. This led to the minimization of the spectral transition measure (STM) [9]

$$D_M(n) = \left\| \sum_{m=-M}^{M} m\mathbf{y}\,(n+m) \right\|^2 , \tag{9}$$

where $2M + 1$ is the location window length.

By inspection of Eq. (8), TRISM is seen to be a normalized measure in relation to the spectral coefficients and to the location window length, whereas, by Eq. (9), STM is found to depend directly on the magnitude of the spectral coefficients. The weighting of the spectral coefficients is seen to be symmetric for TRISM and antisymmetric for STM as shown additionally in Fig. 1. This can be interpreted to implicitly involve symmetric interpolation functions in the evaluation of TRISM and asymmetric ones in the evaluation of STM as illustrated in Fig. 2.

For computational cost evaluation, Eq. (9) can be rearranged as follows

$$D_M(n) = \left\| \sum_{m=1}^{M} m\,[\mathbf{y}\,(n+m) - \mathbf{y}\,(n-m)] \right\|^2 . \tag{10}$$

The operational complexity involved in the evaluation of MINTRISM and STM for a frame, according to Eqs. (8) and (10) is displayed in Table I, where it can be verified that TRISM requires one addition, p multiplications and one division per frame more than STM for the location of event centers, where p is the LP order. But the greater stability of TRISM evaluations allows for a reduction in the number of refinement iterations in comparison.

TABLE I

OPERATIONAL COMPLEXITY PER FRAME FOR TRISM AND STM EVALUATION, WHERE p IS LP ORDER AND $2M + 1$ IS THE LOCATION WINDOW LENGTH.

Operation	TRISM	STM
$+$	$2Mp$	$2Mp - 1$
$\times$	$(M + 2)p$	$(M + 1)p$
$\div$	1	0

III. LOCAL TEMPORAL DECOMPOSITION

The measures presented in Section II locate the internal event centers C_j for $j = 1, 2, \ldots, J-2$. Additionally, endpoint

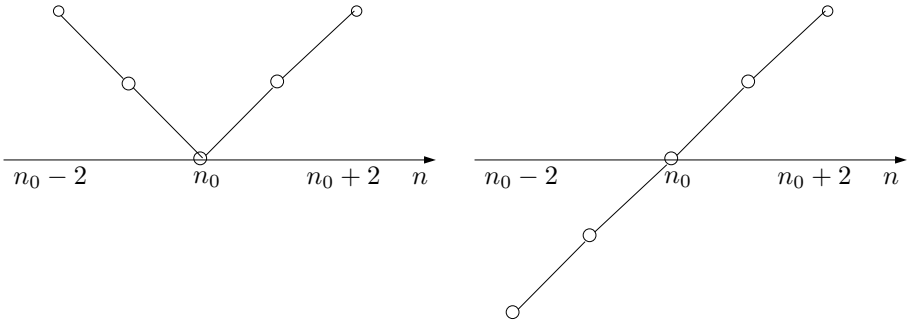

Fig. 1. Location windows for the evaluation of TRISM (left plot) and STM (right plot), illustrated for a five-frame long case.

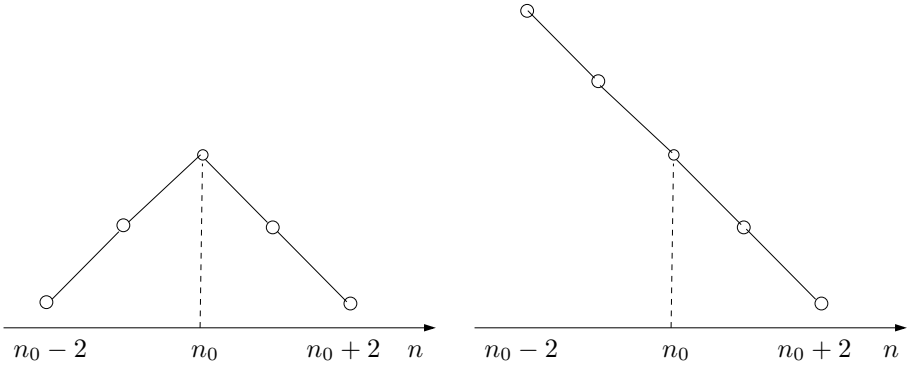

Fig. 2. General shape of event functions involved in the evaluation of TRISM (left plot) and STM (right plot), illustrated for a five-frame long case.

event centers are located at $C_0 = 1$ and $C_{J-1} = N - 1$. The detected event rate is

$$f_e = \frac{J-1}{\sum_{j=1}^{J-1} C_j - C_{j-1}} f_f, \tag{11}$$

where f_f is the frame rate for LP analysis.

Each target vector is initially identified to the original LSF vector at the event center just located, that is,

$$\mathbf{a}_j = \mathbf{y}(C_j) \tag{12}$$

for $j = 0, 1, \ldots, J-1$.

The two event functions, $\boldsymbol{\phi}_j(n)$ and $\boldsymbol{\phi}_{j+1}(n)$, for the current superframe $n = C_j, C_{j+1}, \ldots, C_{j+1} - 1$, are determined as a function of the right-hand target vector $\mathbf{a}_j$ for the previous superframe and the running estimate $\mathbf{a}_{j+1}$ for the right-hand target vector of the current superframe by the optimal procedure outlined in [13], which consists of the solutions to the sets of equations

$$\begin{bmatrix} \mathbf{a}_j^T \mathbf{a}_j & \mathbf{a}_j^T \mathbf{a}_{j+1} \\ \mathbf{a}_j^T \mathbf{a}_{j+1} & \mathbf{a}_{j+1}^T \mathbf{a}_{j+1} \end{bmatrix} \begin{bmatrix} \hat{\phi}_j(n) \\ \hat{\phi}_{j+1}(n) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_j^T \mathbf{y}(n) \\ \mathbf{a}_{j+1}^T \mathbf{y}(n) \end{bmatrix} \tag{13}$$

for $n = C_j, C_j+1, \ldots, C_{j+1}-1$. The estimated event function samples in Eq. (13) are modified, if necessary, to lie in the range from zero to unity, that is,

$$\begin{aligned} \phi_j(n) &= \min\left\{1, \max\left\{0, \hat{\phi}_j(n)\right\}\right\} \\ \phi_{j+1}(n) &= \min\left\{1, \max\left\{0, \hat{\phi}_{j+1}(n)\right\}\right\} \end{aligned} \tag{14}$$

for $n = C_j, C_j + 1, \ldots, C_{j+1} - 1$.

Next, the right-hand target vector $\mathbf{a}_{j+1}$ for the current superframe is reestimated, given the left-hand target vector and the sample values of the event functions in the current superframe, by minimizing the square interpolation error

$$\begin{aligned} \varepsilon_j &= \sum_{n=C_j}^{C_{j+1}-1} \|\mathbf{e}(n)\|^2 \\ &= \sum_{n=C_j}^{C_{j+1}-1} \|\mathbf{y}(n) - \mathbf{a}_j \phi_j(n) - \mathbf{a}_{j+1} \phi_{j+1}(n)\|^2 \end{aligned} \tag{15}$$

Setting the gradient $\frac{\partial \varepsilon_j}{\partial \mathbf{a}_{j+1}} = \mathbf{0}$ and rearranging terms, the new estimate for the right-hand target vector is found to be

$$\mathbf{a}_{j+1} = \frac{\sum_{n=C_j}^{C_{j+1}-1} \mathbf{y}(n)\phi_j(n) - \mathbf{a}_j \sum_{n=C_j}^{C_{j+1}-1} \phi_j(n)\phi_{j+1}(n)}{\sum_{n=C_j}^{C_{j+1}-1} \phi_{j+1}^2(n)}. \tag{16}$$

The LSFs in refined target vectors are tested for stability and made stable by the procedure described in [9] if necessary.

By defining initial event functions as straight-line segments, refinement can be carried out in either order, that is, event functions first or target vector first. Both orders are tested in the experiments described in Section IV.

Refinement is repeated until iteration I such that the relative square interpolation error difference satisfies the inequality

$$\frac{\varepsilon_j^{(I-1)} - \varepsilon_j^{(I)}}{\varepsilon_j^{(I)}} \leq \delta. \tag{17}$$

Also, lower complexity TD algorithms are used that constrain the two event functions in a superframe to be symmetric

unity-complementary [9], [14], that is,

$$\phi_{j+1}(n) = 1 - \phi_j(n) \tag{18}$$

for $n = C_j, C_j + 1, \ldots, C_{j+1} - 1$.

The computational complexity of overall TD analysis is dominated by the second-stage iterative determination of next event target and current and next event functions. Further, the number of iterations in the second stage depends on the method used for event function determination, either the optimal or the symmetric procedures.

IV. Experiments with Temporal Decomposition and Linear Interpolation

Speech spectral envelopes were obtained at a frame rate of 200 Hz as the LSF vector representation that results from tenth-order LP analysis of a 25 ms segment of speech extracted through a Hamming window. The whole set of signals in the test partition of the TIMIT database [15], [16] was resampled at 8 kHz and LP-analyzed as just described, resulting in a total of 1.037 million frames of speech.

Since TRISM is a more stable measure for event location, event rate hardly varies with location window length $2M + 1$. However, by interposing a dead time of M frames after each event detection, frame rate can be controlled when using TRISM. By this procedure event rates may be varied from 12 Hz up to 65 Hz when $M = 1, 2, \ldots, 12$. This same variation in M causes event rates for STM to range from around 12 Hz up to 50 Hz. Different event rates may also be obtained with a fixed window length for all rates by varying the original frame rate [14].

Interpolation error is measured as log spectral distortion (SD) [17] between the original log spectral envelope $10 \log_{10} S_n\left(e^{j\omega}\right)$ and the interpolated log spectral envelope $10 \log_{10} \hat{S}_n\left(e^{j\omega}\right)$ associated with the original LSF vector $\mathbf{y}(n)$ and the interpolated LSF vector $\hat{\mathbf{y}}(n)$, respectively. The log SD is evaluated as the root mean square value $D(n)$ of the difference between these log spectral envelopes over a 1000-point uniform grid on the unit circle.

The minimum relative square interpolation error difference for stopping refinement, defined in Eq. (17), was set to $\delta = 1 \cdot 10^{-4}$, resulting in a mean number of refinement iterations per superframe ranging from 5 to 10.

Two factors may be compared and contrasted by observing Fig. 3, namely, the event location criterion and the order the event functions and the target vectors are reestimated in each refining iteration. For the MINSTM criterion, reestimating the target vector last causes a decrease of about 0.20 dB in distortion along most of the event rate range tested while for MINTRISM the improvement is far from uniform, reaching a maximum of about 0.20 dB at around 30 Hz but giving virtually coincident results below an event rate of around 15 Hz. When target vectors are refined last, the MINTRISM criterion displays a consistent decrease of 0.20 dB over MINSTM for event rates below 30 Hz.

Next, a reference for goodness of fit was sought for the TD algorithms by comparing their overall performance to that of uniform linear interpolation. For the outline of the

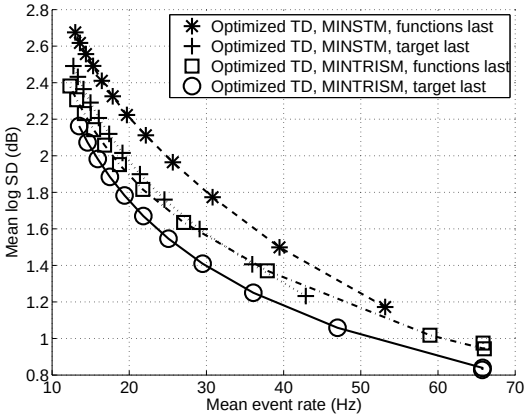

Fig. 3. MINTRISM versus MINSTM criteria for optimal TD with reestimation of event functions and target vector in both orders.

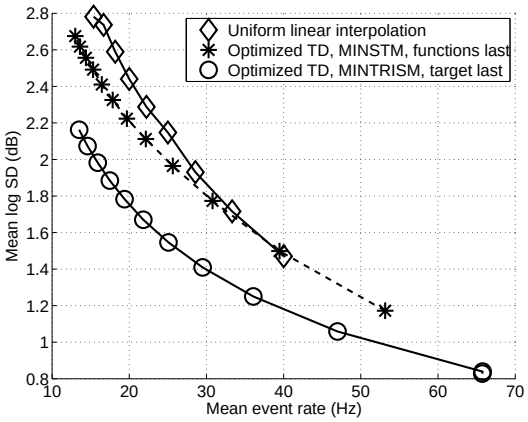

Fig. 4. MINTRISM and MINSTM criteria for optimal TD are compared to uniform linear interpolation. Reestimation of event functions is followed by target vector reestimation for MINTRISM TD and is done the other way around for MINSTM TD.

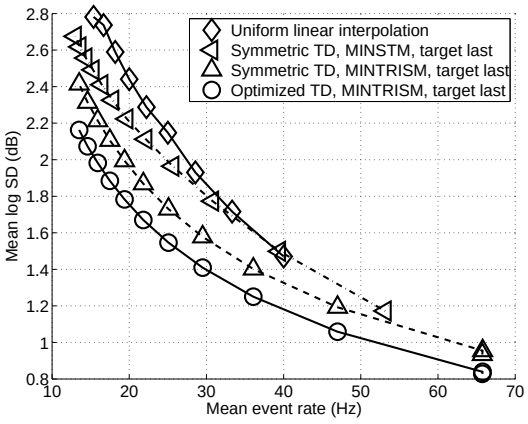

Fig. 5. Symmetric MINTRISM and MINSTM TD are compared to MINTRISM optimal TD and to uniform linear interpolation. Reestimation of event functions is followed by target vector reestimation.

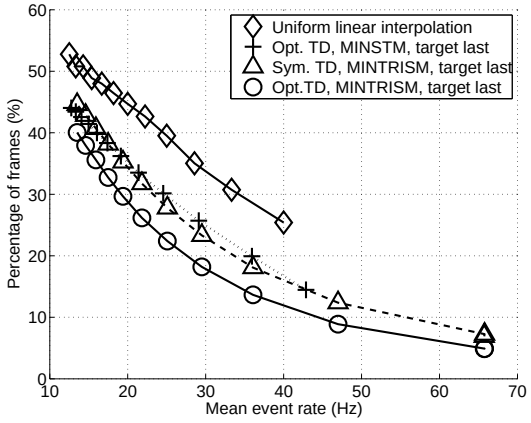

Fig. 6. Percentage of frames n such that 2 dB $< D(n) \leq$ 4 dB for symmetric MINTRISM TD, optimal MINTRISM and MINSTM TD and uniform linear interpolation with reestimation of event functions followed by target vector reestimation.

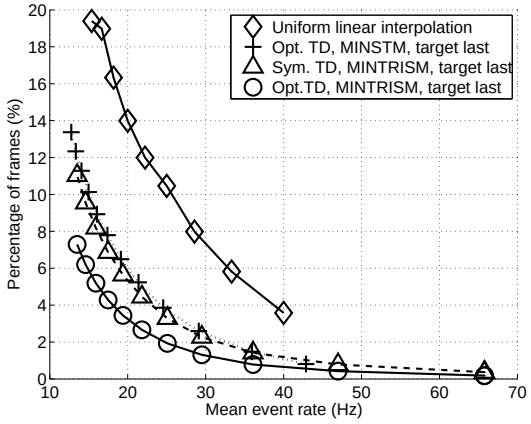

Fig. 7. Percentage of frames n such that $D(n) >$ 4 dB for symmetric MINTRISM TD, optimal MINTRISM and MINSTM TD and uniform linear interpolation with reestimation of event functions followed by target vector reestimation.

global domain of TD performance, just the lowermost and the uppermost curves in Fig. 3 were selected for overlay with the uniform linear interpolation curve in Fig. 4. It can be seen that the distortion for optimal MINTRISM TD with target refinement last is always lower than that of linear interpolation by at least 0.4 dB for mean event rates below 33 Hz. Even the upper distortion bound for optimal TD lies below the linear interpolation distortion curve for all mean event rates below about 36 Hz.

The symmetric low-complexity local TD algorithms are compared to the best optimal TD algorithm and to uniform linear interpolation in Fig. 5. For mean event rates below 30 Hz, the low-complexity MINTRISM TD performance is uniformly 0.2 dB higher in distortion than the optimal algorithm, whose distortion is lower than that of linear interpolation by 0.3 dB at the higher event rates to more than 0.5 dB at the lower rates. On the other hand, the performance of the symmetric MINSTM TD algorithm traverses from the meeting point with linear interpolation at about 37 Hz to a virtual encounter with symmetric MINTRISM TD at about 12.5 Hz. The distribution along the frames of the log SD for LSF interpolation may be better assessed through the analysis of Figs. 6 and 7, which display the percentage of outliers in the range above 2 dB and up to 4 dB and in the range above 4 dB, respectively. As a group, TD algorithms show from half to one-fifth as much percentage of outermost outliers as linear interpolation. Inside the TD group, the behavior of the best symmetric TD algorithm is confined between those of optimal MINTRISM TD and optimal MINSTM TD. The operation of linear interpolation at 33.33 Hz may be taken as acceptable since it is used in low-bit-rate coding [18], [10]. For the same outermost outlier percentage, optimal MINSTM TD operates at a mean event rate of 20 Hz and optimal MINTRISM TD operates at around 15 Hz as shown in Fig. 7. This means over two times a compression ratio for MINTRISM TD over linear interpolation. These event rates include on average 10 and 13 frames per superframe, respectively. In addition, low-complexity symmetric MINTRISM TD operates at virtually the same event rate as optimal MINSTM TD.

V. Conclusion

Variable-rate sampling and interpolation of LSF tracks for speech signals has been analyzed and tested, using uniform linear interpolation as a baseline for comparison. The proposed algorithm features low algorithmic delay due to sequential event location. Events are localized by the first stage of the algorithm using the proposed minimum triangular interpolation spectral measure (MINTRISM) criterion. The mean realized event rate under MINTRISM is the least sensitive to location window length among global and local TD criteria. Refining target vectors after event functions improves the spectral match, particularly at higher event rates, but the order of refinement is immaterial below a mean event rate of 20 Hz.

Over a mean event rate range from 12 Hz up to 35 Hz, TRISM performs better than STM by 0.2 dB in log SD. A lower complexity version of MINTRISM TD constrains the two event functions in a superframe to be symmetric unity-complementary and performs between MINSTM and MINTRISM TD. They can interpolate a maximum of 10, 11 and 13 frames between adjacent events, for a uniform frame rate of 200 Hz, within the interpolation distortion of standard low-bit-rate speech coders.

Acknowledgment

The author is grateful to the Associate Editors and anonymous reviewers, whose suggestions and corrections have substantially enhanced this paper.

References

[1] F. Itakura, “Line spectral representation of linear predictor coefficients of speech signals,” *J. Acoust. Soc. Am.*, vol. 57, no. S1, p. S35, Apr. 1975.

[2] M. Arjona Ramírez and M. Minami, “Technology and standards for low-bit-rate vocoding methods,” in *The Handbook of Computer Networks*, H. Bidgoli, Ed. New York: Wiley, 2008, vol. 2, pp. 447–467.

[3] F. K. Soong and B.-H. Juang, “Line spectrum pair (LSP) and speech data compression,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, San Diego, 1984, pp. 1.10.1–1.10.4.

[4] B. S. Atal, “Efficient coding of LPC parameters by temporal decomposition,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, Boston, 1983, pp. 81–84.

[5] M. Taylor and F. Bimbot, “Temporal decomposition for the initialization of a HMM isolated word-recognizer,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, San Francisco, 1992, pp. 369–372.

[6] C. R. N. Athaudage, A. B. Bradley, and M. Lech, "Model-based speech signal coding using optimized temporal decomposition for storage and broadcasting applications," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1016–1026, Mar. 2003.

[7] R. Moldover and A. Kain, "Compression of line spectral frequency parameters with asynchronous interpolation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Taipei, 2009, pp. 3789–3792.

[8] A. C. R. Nandasena and M. Akagi, "Spectral stability based event localizing temporal decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, Seattle, 1998, pp. 957–960.

[9] S.-J. Kim and Y.-H. Oh, "Efficient quantisation method for LSF parameters based on restricted temporal decomposition," *Electron. Lett.*, vol. 35, no. 12, pp. 962–964, June 1999.

[10] K. K. Paliwal, "Interpolation properties of linear prediction parametric representations," in *Proc. European Conference on Speech Communication and Technology*, Madrid, 1995, pp. 1029–1032.

[11] C. R. Ferreira, A. Alcaim, and R. C. de Lamare, "Modified interpolation of LSFs based on optimization of distortion measures," in *Proc. SBrT/IEEE International Telecommunications Symposium*, vol. 1. Fortaleza, Brazil: IEEE, 2006, pp. 777–782.

[12] R. C. de Lamare and A. Alcaim, "Strategies to improve the performance of very low bit rate speech coders and application to a 1.2 kb/s codec," *IEE Proc.-Vis. Image Signal Process.*, vol. 152, no. 1, pp. 74–86, Feb. 2005.

[13] C. R. N. Athaudage and M. Lech, "On optimal modelling of speech spectral transitions," in *Proc. Information, Communications and Signal Processing 2003 and the Fourth Pacific Rim Conference on Multimedia*, vol. 3, Singapore, 2003, pp. 1330–1334.

[14] M. Arjona Ramírez and V. O. P. Machado, "Temporal decomposition of parameter tracks for speech coding," in *Proc. of the International Workshop on Telecommunications*, Santa Rita do Sapucaí, 2007, pp. 199–203.

[15] W. M. Fisher, G. R. Doddington, and K. M. Goudie-Marshall, "The DARPA speech recognition research database: specifications and status," in *Proc. DARPA Speech Recognition Workshop*, vol. 1, Feb. 1986, pp. 93–99.

[16] 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. [Online]. Available: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1

[17] M. Arjona Ramírez and M. Minami, "Split-order linear prediction for segmentation and harmonic spectral modeling," *IEEE Signal Processing Lett.*, vol. 13, no. 4, pp. 244–247, Apr. 2006.

[18] ITU-T, "*Dual rate speech coder for multimedia applications transmitting at 5.3 and 6.3 kbit/s*," Recommendation G.723.1, Geneva, Mar. 1996.

Miguel Arjona Ramírez 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 Inform´atica, 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.