Correspondence

LPC Interpolation by Approximation of the Sample Autocorrelation Function

Jan S. Erkelens and Piet M. T. Broersen

***Abstract*—Conventionally, the energy of analysis frames is not taken into account for linear prediction (LPC) interpolation. Incorporating the frame energy improves the subjective quality of interpolation, but increases the spectral distortion (SD). The main reason for this discrepancy is that the outliers are increased in low energy parts of segments with rapid changes in energy. The energy is most naturally combined with a normalized autocorrelation representation.**

***Index Terms*—LPC interpolation, speech coding.**

I. Introduction

Speech signals are considered mostly stationary over relatively short time intervals of about 25 ms in speech coding. In such intervals, linear prediction-based (LPC-based) speech coders describe the speech spectral envelope with an autoregressive model. Models from consecutive frames can be very different in transition segments. Updating the LPC model more frequently could be used to follow changes in the spectral properties, but this would increase the bit rate. A more efficient solution is interpolation of the LPC models of consecutive analysis frames. Proper interpolation has a smoothing effect, which turns out to be beneficial for speech quality.

For speech synthesis a stable LPC filter is required. The LPC parameters themselves, i.e., the coefficients of the all-pole LPC synthesis filter, do not ensure stability when interpolated. Representations that satisfy the stability requirement are autocorrelation representations, reflection coefficients (RC's), log area ratios (LAR's), arcsine of the reflection coefficients (ASRC's) and line spectral pairs (LSP's) [1]–[4].

Some other studies on interpolation of the LPC model are reported in the literature [5]–[10]. Most of these studies compared the interpolation behavior of different representations by means of experimental results on speech data. In this correspondence, we will follow a more theoretical approach. A logical definition of an optimal interpolation method is that the interpolated model for a subframe is as close as possible to the true subframe model, i.e., the model that would be obtained from LPC analysis for that subframe. From this definition a new interpolation method is developed in Section II, which uses the energy of the analysis frames. Experimental results will be given in Section III. Conclusions follow in Section IV.

Manuscript received November 27, 1995; revised November 20, 1997. This work was supported by the Dutch Technology Foundation under Grant DTN11.2436. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Douglas D. O'Shaughnessy.

J. S. Erkelens is with the Faculty of Information Technology and Systems, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: j.s.erkelens@et.tudelft.nl).

P. M. T. Broersen is with the Faculty of Applied Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands (e-mail: broersen@tn.tudelft.nl).

Publisher Item Identifier S 1063-6676(98)07790-6.

II. Interpolation by Approximation of the Sample Autocorrelation Function

The purpose of interpolation of the spectral model is to avoid the increase in bit rate resulting from an increased sampling rate of the LPC parameters. Consider Fig. 1, where different frames and models are defined: in LPC analysis of speech, models are obtained from short time segments, called *analysis frames*. Suppose a model of the signal is wanted more often, to follow changes in the spectral properties of the signal. Estimating the LPC model more frequently in the *intermediate frames* would lead to an increase in the bit rate of a speech coding system. *Regular models* are defined as the models obtained from the analysis frames; *reference models* are obtained from the intermediate frames. The increase in bit rate can be avoided by interpolation of the regular models of the analysis frames. An obvious requirement for an interpolation method is that the interpolated models are as close as possible to the reference models from the intermediate frames.

This definition of optimal interpolation is logical because an intermediate frame could have been a regular analysis frame, if the coding session had been started somewhat earlier or later. In this section, a new interpolation method will be developed, maintaining the above definition. The following conditions are assumed, although they are not essential. The signal is stationary within the analysis frames; the regular analysis frames are bordering, i.e., are not overlapping, and no tapered data window is applied for LPC analysis. It will be shown that the given definition of optimal interpolation requires the use of the frame energy in the interpolation. Furthermore, it will be shown that any interpolation method that assumes stationarity within an analysis frame will not deal well with the low energy part of segments with rapidly changing energy (for example, onsets), if these sudden changes occur within a frame. In this case, the assumption that the signal within the frame is stationary is not valid and it would actually be better to adapt the location of the analysis-frame borders to the signal characteristics, as is used in for example *phonetic segmentation* [11]. In this correspondence, only fixed borders will be considered.

In Fig. 2, an unvoiced/voiced transition is shown. Consider two analysis frames in this figure, located from 0 ms to 20 ms and from 20 ms to 40 ms, respectively. The first analysis frame is entirely unvoiced, the second one entirely voiced. The actual energy of each analysis frame has no influence on the LPC parameters of that frame: if one of the analysis frames is multiplied by a constant (changing the energy of the frame), the same LPC model is obtained. However, the situation is different for reference models of intermediate frames. For example, consider an intermediate frame centered at the boundary, located from 10 ms. to 30 ms. Suppose the energy of the first analysis frame is increased, say by a factor 100, while the energy of the second analysis frame is not changed. In this case, a completely different model will be obtained from the intermediate frame, because the energy of the first half is altered. However, the model obtained by the interpolation of one of the conventional representations will remain the same. This is not according to our definition of optimal interpolation: a good interpolation method must aim at approximating the reference model obtained from the intermediate frames. The main conclusion is that the interpolation should depend on the energy of the analysis frames, because the reference models are influenced by these energies [10].

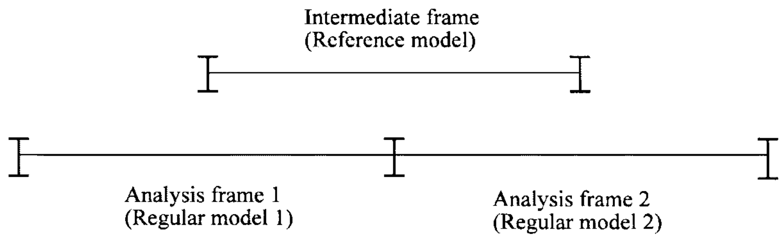

Fig. 1. Definition of different frames used in the text. The regular LPC models are obtained from the analysis frames. Interpolation of these models should approximate the reference models from the intermediate frames.

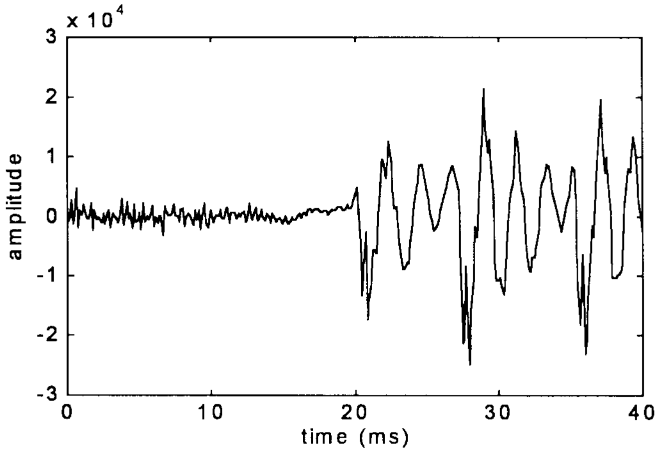

Fig. 2. Unvoiced-voiced transition; including the energy biases the interpolation toward the frame with highest energy.

One way to incorporate the energies of analysis frames is to multiply the coefficients of an LPC representation with the energy of the corresponding analysis frame and to interpolate this energy weighted representation. For example, a vocoding algorithm is proposed in [12] that puts great effort into accurately coding the time envelope of the speech signal, because the energy contains important perceptual information. In that algorithm the frame energy is used in the LPC interpolation, in connection with LSP's. It is mentioned that energy weighting biases the interpolation toward the frame with highest energy; improving the performance of the coder at rapid onsets. There is, however, a more natural way to incorporate the frame energy in the interpolation. Autoregressive models describe the autocorrelation of a signal in the time domain and the spectral envelope in the frequency domain. Since autoregressive estimation methods use the sample autocorrelation function of the data to obtain LPC parameters, the interpolation must aim at giving the best approximation of the sample autocorrelation function of intermediate frames. The sample autocorrelation function of analysis frames contains the energy of these frames. Since intermediate frames contain a part of two consecutive analysis frames, a reconstruction of the sample autocorrelation function of intermediate frames can be obtained by interpolation of the sample autocorrelation functions of the analysis frames (if stationarity is assumed within the analysis frames). An autocorrelation representation of the LPC model should be used for interpolation, but one that includes the analysis-frame energy. Autocorrelation representations have been used for interpolation in the past (e.g., in [4] and [5]), but not one that uses the actual analysis-frame energy.

Autocorrelation representations can be normalized in different ways. To illustrate this, consider Fig. 3, where an all-pole filter $1/A(z)$ is excited by a white excitation $e(n)$ to give a signal $x(n)$. Let σ_e^2 and σ_x^2 be the power of $e(n)$ and $x(n)$, respectively. The following relationship exists between these powers:

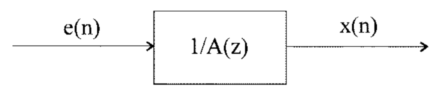

Fig. 3. The autocorrelation of the output $x(n)$ of an all-pole filter $1/A(z)$, with a white-noise signal $e(n)$ as the input, can be normalized in different ways (see text).

$$\sigma_x^2 = \frac{\sigma_e^2}{\prod_{i=1}^{p} \left\{1 - k_i^2\right\}}$$

where p is the order of the all-pole filter of Fig. 3 and the k_i are the reflection coefficients associated with it. Autocorrelation representations that are of interest for the problem of interpolation are defined as follows. The *autocorrelation coefficients* (ACR's) describe the autocorrelation of $x(n)$ with σ_x^2 normalized to one. The *autocovariance coefficients* (ACV's) describe the autocorrelation of $x(n)$ with the *input* variance σ_e^2 normalized to 1. Interpolation of ACR has been used for example in codebook design [13]. ACR's and ACV's have σ_x^2 and σ_e^2 normalized to one, respectively, and therefore *do not use the actual frame energy*, but we are looking for a normalization that does use it. The *autocorrelation function* (ACF) describes the autocorrelation of $x(n)$ when σ_x^2 is made equal to the actual frame energy. Both ACR's and ACV's can be computed directly from the LPC parameters [4]. ACR's can also be computed from the ACF by normalizing with the frame energy. At the decoder, the ACF can be recovered by multiplying ACR with the frame energy. Because the frame energy is usually transmitted to the decoder, this does not cause an increase in bit rate.

As an illustration, we averaged the models of the previously defined analysis frames in Fig. 2 by means of ACR, ACV, and ACF, respectively (Averaging is interpolation with equal weights). When the averaged models were compared with the reference model of the intermediate frame, the spectral distortion (SD) measure was 5.55 dB, 3.80 dB, and 1.20 dB, respectively. Clearly, the actual energy of analysis frames must be taken into account. When we decreased the energy of the first analysis frame (increasing the relative energy differences), ACR's and ACV's performed even worse. Only when the frames are scaled such that their energies are equal, ACR's and the ACF give identical results, because then there is no energy difference.

The ACF is a new interpolation method that takes into account the actual energy of analysis frames. The frame energy can be used in connection with other representations. However, only with the ACF is the frame energy incorporated in a natural way.

When energy weighting is applied for interpolation of a representation other than ACR, a capital E will be added to the abbreviation of that representation. For example: energy-weighted LSP interpolation will be referred to as *ELSP interpolation*. Energy-weighted interpolation is performed by multiplying the coefficients of a representation with the frame energy and interpolating these

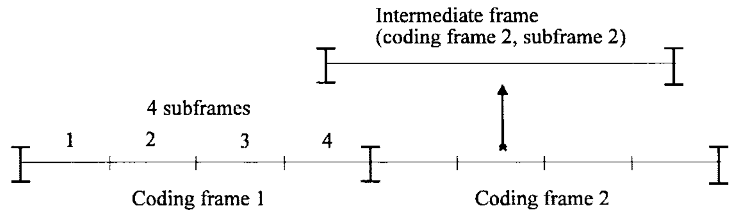

Fig. 4. Definition of frames used in the experiments.

TABLE I
SD AND NUMBER OF OUTLIERS FOR INTERPOLATION IN TEN SPEECH SENTENCES. THE BURG METHOD WAS USED FOR ANALYSIS IN NONOVERLAPPING ANALYSIS FRAMES OF LENGTH 25 ms. NO TAPERED DATA WINDOW WAS APPLIED FOR THE LEFT-HAND SIDE. A HAMMING WINDOW WAS APPLIED FOR THE RIGHT-HAND SIDE.

Burg method, no window, 25 ms				Burg method, Hamming, 25 ms			
interp. method	SD (dB)	outliers (%)		interp. method	SD (dB)	outliers(%)	
		2-4 dB	> 4 dB			2-4 dB	> 4 dB
ACR	1.72	17.8	8.9	ACR	1.93	26.5	9.5
ACV	1.55	17.0	6.6	ACV	1.77	26.6	6.7
ACF	1.77	16.6	10.2	ACF	2.05	27.3	11.5
LSP	1.44	16.3	4.5	LSP	1.62	23.3	4.8
ELSP	1.80	14.2	10.9	ELSP	2.08	24.5	12.2
LAR	1.63	17.3	7.9	LAR	1.85	25.1	8.5
ELAR	1.88	19.2	11.0	ELAR	2.16	28.2	12.8

energy-weighted coefficients linearly. In other words, the interpolation weighting factors become dependent on the frame energy.

Although we have tacitly assumed bordering frames and no windowing, it is clear that the frame energy should also be taken into account when overlapping or windowed frames are used, because large differences in the energy of consecutive frames must be properly treated also under these analysis conditions.

III. EXPERIMENTAL RESULTS

In the previous section, it was argued that the energy of analysis frames should be incorporated in the interpolation, and it was shown that this can be done in a natural way with energy-weighted ACR's (ACF interpolation), if the signal is stationary in an analysis frame. Experimental results are presented in this section to test the new interpolation method, with different analysis conditions. In the previous section, we focused on the subject of interpolation for nonoverlapping nonwindowed analysis frames. In speech coders, these assumptions are not always satisfied. Often, the autocorrelation method is used to obtain LPC models from overlapping windowed analysis frames. Therefore, experimental results will be given also under these analysis conditions.

In the experiments, we used ten sentences from the TIMIT data base, downsampled to an 8 kHz sample frequency. This set contained five male and five female speakers. In Fig. 4, it is shown how the frames are defined for these experiments. LPC analysis was performed every 25 ms (200 samples). In some of the experiments, the analysis frames are taken to be longer than 25 ms and are therefore overlapping in these cases. Associated with each analysis frame is a *coding frame*, which consists of four subframes (Figs. 1 and 4). The coding frames do not overlap. The analysis frames are centered around the middle of the coding frames. The regular models from the analysis frames are interpolated with the different representations and the resulting interpolated models are compared to the *reference models* in terms of SD and number of outliers. These reference models are obtained from the *intermediate frames*, which are centered around the middle of the subframes (Fig. 4). The LPC residual signals in the subframes were computed with the reference models. These residuals will be used as the input for the interpolated filters in the listening experiments. Hence, for all interpolation methods, the same excitation is used for synthesis. This ensures that all objective and subjective differences between the interpolation methods are due to inaccuracies in the interpolation only.

Objective results are shown in Tables I and II. All results are obtained from a total of 3600 intermediate frames. In Table I, results are presented for the Burg method, with and without windowing. The left-hand side of the table presents results when no window is applied; the right-hand side shows results when a Hamming window is applied. In Table II results are given for the autocorrelation method with a Hamming window for different degrees of overlap. In the left-hand part of the table, results are given for analysis frames of length 30 ms; for the right-hand part, analysis frames are of length 37.5 ms. For all experiments, LPC analysis was performed every 25 ms. The results show that using the frame energy in the interpolation leads to a higher average SD and number of outliers. *This is in contrast with the subjective test* to be described next: ACF interpolation subjectively outperforms LSP interpolation.

Different conventional interpolation methods have been compared in the literature, and those based on LSP perform very well, both objectively and subjectively [6]–[9]. Our results show that the ACF performs objectively the best of all energy-weighted representations. To decide whether energy weighting improves the subjective quality of interpolation, a direct confrontation between the ACF and LSP was carried out by means of an AB-preference test with seven listeners. For each of the ten sentences, the reconstructions obtained with ACF

TABLE II
SD AND NUMBER OF OUTLIERS FOR INTERPOLATION IN TEN SPEECH SENTENCES. LPC ANALYSIS WAS PERFORMED EVERY 25 ms WITH THE AUTOCORRELATION METHOD IN OVERLAPPING ANALYSIS FRAMES OF LENGTH 30 MS FOR THE LEFT-HAND SIDE AND 37.5 MS FOR THE RIGHT-HAND SIDE. A HAMMING WINDOW WAS APPLIED

Autocorrelation method, Hamming, 30 ms				Autocorrelation method, Hamming, 37.5 ms			
interp. method	SD (dB)	outliers (%)		interp. method	SD (dB)	outliers(%)	
		2-4 dB	> 4 dB			2-4 dB	> 4 dB
ACR	1.83	24.3	8.0	ACR	1.63	20.0	6.5
ACV	1.69	24.9	5.6	ACV	1.47	18.6	4.2
ACF	1.95	25.8	10.0	ACF	1.72	20.2	8.4
LSP	1.55	21.7	4.1	LSP	1.35	17.2	2.5
ELSP	1.98	23.3	11.0	ELSP	1.75	18.2	9.3
LAR	1.78	23.5	7.6	LAR	1.57	18.4	6.1
ELAR	2.07	26.5	11.8	ELAR	1.86	22.1	10.0

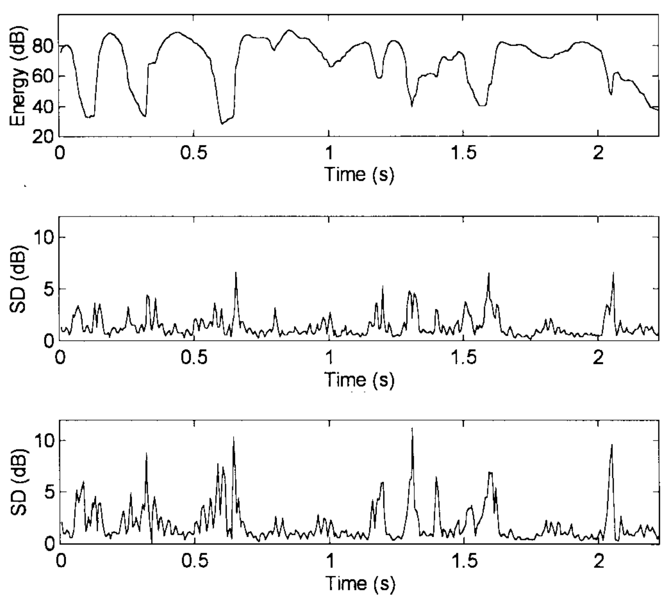

Fig. 5. (a) Energy of a speech sentence. (b) SD for LSP interpolation. (c) SD for ELSP interpolation. Objectively, interpolation does not work well in the low energy parts of segments with rapidly changing energy.

and LSP interpolation were played in random order, and the listeners could listen up to three times to the reconstructions before giving their preference. Preference was given to ACF in 61.4% of the cases. This interesting result shows that the new interpolation method is indeed better in a subjective sense, contradicting the objective results. The main reason for this discrepancy is the higher number of outliers in low-energy parts of the signal, as will be explained below. These outliers do not influence the subjective quality in a negative way. This means that a low SD is a *sufficient* condition for a good subjective quality, but not always a *necessary* one. This has also been observed in the context of quantization [14].

SD and the number of outliers have been used successfully in quantization as an objective measure of the quality of the quantizer. It is known approximately when a quantizer achieves transparent quality [15]: when the average SD is smaller than 1 dB, there are less than 2% outliers in the range 2–4 dB and there are no outliers larger than 4 dB. It has been observed [5], [7], [8] that the number of outliers for *interpolation* is much larger than what is allowed for *quantization*. This large number of outliers also influences the average value of SD, which is larger than 1 dB for interpolation. The large number of outliers can be explained. The energy of the intermediate frames of one speech sentence is shown in Fig. 5(a), expressed in decibels. In Fig. 5(b) SD is shown for LSP interpolation. It is clear that the largest errors are made in segments where large changes in energy occur, such as silence-speech and unvoiced-voiced transitions. Such a sudden change in energy will generally not occur exactly at a frame boundary. Suppose that in Fig. 2 the transition occurs somewhere in the middle of the second analysis frame. Then an intermediate frame located in this figure from 5 ms to 25 ms would be entirely unvoiced. Interpolation will not give a good model for this intermediate frame because the model from the second analysis frame (which is then mixed unvoiced-voiced) does not fit well to the model from the first, unvoiced, analysis frame. Using the energy biases the interpolation toward the frame with highest energy, and will give a less accurate approximation for the models for the low energy part of the transition. This can be clearly seen in Fig. 5(c), where SD is shown for ELSP interpolation in the same sentence. The listening experiment showed that the larger number of outliers in low energy parts is not decreasing the subjective quality. Energy-weighting gives a better model for the high energy part of transitions. It has been reported in [12] that energy-weighted interpolation improves the performance of a coder at rapid onsets. The larger number of outliers in low-energy parts does, however, influence the objective results in terms of SD. It is therefore difficult to make a fair comparison between conventional interpolation and energy-weighted interpolation exclusively on the basis of SD and number of outliers.

Some other interesting points can be noticed in the tables. Theoretically, the Burg method does not need a tapered data window. Table I shows that application of a tapered data window decreases the objective performance for the Burg method. The autocorrelation method does need a tapered data window and Table II shows that increasing the overlap of analysis frames increases the performance. This was also observed for the Burg method. A more detailed analysis of the autocorrelation method and the role of windowing is presented elsewhere [16].

The reflection coefficient based representations (RC, LAR, and ASRC) perform worse than the other representations. It can be shown that the coefficients of these representations may have strong nonlinear dependencies [10], [17], which are undesirable for linear interpolation.

IV. Conclusion

In this work, it is shown that the actual energy of analysis frames should be taken into account for interpolation. The required approximation of the sample autocorrelation function can be implemented by multiplying the autocorrelation coefficients with the frame energy and interpolating this function (ACF interpolation). ACF interpolation outperformed LSP interpolation in a subjective test, contrasting the objective results.

The main reason for the discrepancy between subjective and objective results is that the largest outliers occur in low energy parts of segments with rapidly changing energy and it turned out that these do not have much influence on the subjective quality.

References

[1] F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, p. S35, 1975.

[2] V. R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 309–321, 1975.

[3] A. H. Gray and J. D. Markel, "Quantization and bit allocation in speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 459–473, 1976.

[4] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, pp. 637–655, 1971.

[5] B. S. Atal, R. V. Cox, and P. Kroon, "Spectral quantization and interpolation for CELP coders," in *Proc. Int. Conf. ICASSP*, 1989, pp. 69–72.

[6] T. Umezaki and F. Itakura, "Analysis of time fluctuating characteristics of linear predictive coefficients," in *Proc. Int. Conf. ICASSP*, 1986, pp. 1257–1260.

[7] M. Yong, "A new LPC interpolation technique for CELP coders," *IEEE Trans. Commun.*, vol. 42, pp. 34–38, 1994.

[8] K. K. Paliwal, "Interpolation properties of linear prediction parametric representations," in *Proc. Int. Conf. EUROSPEECH*, 1995, pp. 1029–1032.

[9] H. B. Choi, W. T. K. Wong, B. M. G. Cheetham, and C. C. Goodyear, "Interpolation of spectral information for low bit rate speech coding," in *Proc. Int. Conf. EUROSPEECH*, 1995, pp. 1033–1036.

[10] J. S. Erkelens and P. M. T. Broersen, "Interpolation of autoregressive processes at discontinuities: Application to LPC based speech coding," in *Proc. Int. Conf. EUSIPCO*, 1994, pp. 935–938.

[11] R. Hagen, E. Paksoy, and A. Gersho, "Variable rate spectral quantization for phonetically classified CELP coding," in *Proc. Int. Conf. ICASSP*, 1995, pp. 748–751.

[12] I. A. Atkinson, A. M. Kondoz, and B. G. Evans, "1.6 kbit/s LP vocoder using time envelope," *Electron. Lett.*, vol. 31, pp. 517–519, 1995.

[13] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," in *Proc. IEEE*, 1985, vol. 73, pp. 1551–1588.

[14] J. S. Erkelens and P. M. T. Broersen, "Quantization of the LPC model with the reconstruction error distortion measure," in *Proc. Int. Conf. EUSIPCO*, 1996, pp. 1677–1680.

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

[16] J. S. Erkelens and P. M. T. Broersen, "Bias propagation in the autocorrelation method of linear prediction," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 116–119, 1997.

[17] J. S. Erkelens, *Autoregressive Modeling for Speech Coding: Estimation, Interpolation and Quantization*. Delft, The Netherlands: Delft Univ. Press, 1996.

An Improved (Auto:I, LSP:T) Constrained Iterative Speech Enhancement for Colored Noise Environments

Bryan L. Pellom and John H. L. Hansen

***Abstract*—In this correspondence we illustrate how the (Auto:I, LSP:T) constrained iterative speech enhancement algorithm can be extended to provide improved performance in colored noise environments. The modified algorithm, referred to here as *noise adaptive* (*Auto:I, LSP:T*), operates on subbanded signal components in which the terminating iteration is adjusted based on the *a posteriori* estimate of the signal-to-noise ratio (SNR) in each signal subband. The enhanced speech is formulated as a combined estimate from individual signal subband estimators. The algorithm is shown to improve objective speech quality in additive noise environments over the traditional constrained iterative (Auto:I, LSP:T) enhancement formulation.**

I. Introduction

There are numerous areas where it is necessary to enhance the quality of speech that has been degraded by background distortion. Some of these environments include aircraft cockpits, automobile interiors for hands-free cellular, and voice communications using mobile telephone. Speech enhancement under these conditions can be considered successful if it i) suppresses perceptual background noise and ii) either preserves or enhances perceived speech quality. As voice technology continues to mature, greater interest and demand is placed on using voice-based speech algorithms in diverse, adverse, environmental conditions. It is suggested that the success of advancing speech research in the fields of speaker verification, language identification, and automatic speech recognition could be improved by incorporating front-end speech enhancement algorithms [1].

A number of speech enhancement algorithms have been proposed in the past. A survey can be found in [2], as well as an overview of statistical based approaches in [3]. Several enhancement approaches have been proposed using improved signal-to-noise ratio (SNR) characterization [4], linear and nonlinear spectral subtraction [5], [6], and Wiener filtering [7]. Traditional speech enhancement methods are based on optimizing mathematical criteria, which in general are not always well correlated with speech perception. Several recent methods have also considered auditory processing information [8], [9], and constrained iterative methods using various levels of speech class knowledge [10]–[12].

In this study, we focus on an extension to a previously proposed constrained iterative speech enhancement algorithm termed (Auto:I, LSP:T)[1] [10] (described briefly in Section II). Basically, this method employs spectral constraints on the input speech feature sequence across time and iterations to ensure more natural

Manuscript received February 26, 1997; revised February 26, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jean-Claude Junqua.

The authors are with the Department of Electrical Engineering, Robust Speech Processing Laboratory, Duke University, Durham, NC 27708-0291 USA (e-mail: jhlh@ee.duke.edu).

Publisher Item Identifier S 1063-6676(98)07793-1.

[1] The term (Auto:I, LSP:T) formulated in [10] is derived from the notion that spectral constraints are applied across iterations (I) to the speech autocorrelation lags as well as across time (T) to the speech line spectrum pair (LSP) parameters. For simplicity, (Auto:I, LSP:T) will be referred to as *Auto-LSP* throughout this work.