small. Although increased training data may be helpful, in the speaker dependent Mandarin syllable recognition problem, a limited database will probably still be a normal situation for some period of time in the future.

VII. Conclusion

A new approach is proposed in this correspondence to obtain more elaborate initial models covering characteristics of different tones. Improved state transition topologies are also found to achieve better performance compared with the simple left-to-right model with two transitions. A threshold decision approach is further developed to improve the performance of BS recognition for the syllables with the neutral tone. The test results on everyday Chinese show that a total error rate reduction on the order of 20% in the top 1 rate can be obtained when all the concepts are properly integrated. Although the techniques here are proposed specially for recognition of Mandarin base syllables considering the effect of tones, it is certainly believed that similar concepts are potentially applicable to solve similar problems in speech recognition in other languages.

References

[1] R. He, Ed., *Guoyurbao Tzdian (Mandarin Chinese Daily Dictionary)*. Taipei, R.O.C.: Guoyurbao, 1976.
[2] L.-S. Lee, C.-Y. Tseng, and M. Ouh-Young, "The synthesis rules in a Chinese text-to speech system," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 1309–1320, Sept. 1989.
[3] Y. R. Chao, *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Berkeley Press, 1968.
[4] W. J. Yang, J. C. Lee, Y. C. Chang, and H. C. Wang, "Hidden Markov model for mandarin lexical tone recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 988–992, July 1988.
[5] F.-H. Liu, Y. Lee, and L. S. Lee, "A Direct-concatenation approach to train hidden markov models to recognize the highly confusing mandarin syllables with very limited training data," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 113–119, Jan. 1993.
[6] L.-S. Lee *et al.*, "Golden mandarin (I)—A real-time mandarin speech dictation machine for chinese language with very large vocabulary," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 2, pp. 158–179, Apr. 1993.
[7] B.-H. Juang and L. R. Rabiner, "Mixture autoregressive hidden Markov models for speech signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 6, pp. 1404–1413, Dec. 1985.
[8] X. Huang *et al.*, "The SPHNIX-II speech recognition system: An overview," *Comput. Speech Language*, pp. 137–148, Feb. 1993.

Linear Prediction of the One-Sided Autocorrelation Sequence for Noisy Speech Recognition

Javier Hernando and Climent Nadeu

***Abstract*— The aim of this correspondence is to present a robust representation of speech based on AR modeling of the causal part of the autocorrelation sequence. In noisy speech recognition, this new representation achieves better results than several other related techniques.**

I. Introduction

Linear predictive coding (LPC) [1] is a spectral estimation technique widely used in speech processing and, particularly, in speech recognition. However, the conventional LPC technique, which is equivalent to AR modeling of the signal $x(n)$, is known to be very sensitive to the presence of background noise. This fact leads to poor recognition rates when this technique is used in speech recognition under noisy conditions, even if only a moderate level of contamination is present in the speech signal. Similar results are obtained with the well-known mel-cepstrum technique [2]. This explains why some of the main attempts to combat the noise problem consist of finding novel acoustic representations that are more resistant to noise corruption than traditional parameterization techniques.

Linear prediction of the autocorrelation sequence has been the common approach to several robust spectral estimation methods for noisy signals presented in the past. For speech recognition, Mansour and Juang [3] proposed the short-time modified coherence (SMC) as a robust representation of speech based on that approach. On the other hand, Cadzow [4] introduced the use of an overdetermined set of Yule–Walker equations for robust modeling of time series. Although Cadzow applies linear prediction to the signal, his method can also be interpreted as performing linear prediction in the autocorrelation domain. Both methods rely, either explicitly or implicitly, on the fact that the autocorrelation sequence is less affected by broadband noise than the signal itself, especially at high lag indices.

In this work, we consider the one-sided or causal part of the autocorrelation sequence and its mathematical properties. As this sequence shares its poles with the signal $x(n)$, it provides a good starting point for LPC modeling. In this way, the new one-sided autocorrelation LPC (OSALPC) method appears as a straightforward result of the approach [5]. In addition, it is closely related to the SMC representation and Cadzow's method. All of them can be interpreted as AR modeling of either a spectral function named "envelope" or its square. This interpretation, which is based on the properties of the one-sided autocorrelation, provides more insight into the various methods. In this correspondence, their performance in noisy speech recognition is compared. The optimum model order and cepstral liftering have also been investigated in noisy conditions. The simulation results show that OSALPC outperforms the other techniques in severe noisy conditions and obtains similar scores for moderate or high SNR.

Manuscript received February 14, 1995; revised November 9, 1995. This work was supported by Grant nos. TIC-92-0800-C05/04 and TIC-92-1026-C02/02. The associate editor coordinating the review of this paper and approving it for publication was Dr. Kuldip K. Paliwal.

The authors are with the Department of Signal Theory and Communications, Polytechnical University of Catalonia, Barcelona, Spain.

Publisher Item Identifier S 1063-6676(97)00766-9.

This correspondence is organized in the following way. In Section II, the OSALPC technique is introduced, and its relationship with the conventional LPC approach and the other parameterizations based on AR modeling in the autocorrelation domain is discussed. Section III reports the application of all those parameterization techniques to an isolated word multispeaker recognition task, using the HMM approach, in order to compare their performances in the presence of additive white noise. Finally, some conclusions are summarized in Section IV.

II. AR Modeling in the Autocorrelation Domain

From the autocorrelation sequence $R(m)$, we define the one-sided (causal part of the) autocorrelation (OSA) sequence in the following way:

$$R^+(m) = \begin{cases} R(m) & m > 0 \\ \dfrac{R(0)}{2} & m = 0 \\ 0 & m < 0 \end{cases} \tag{1}$$

Its Fourier transform is the complex "spectrum"

$$S^+(\omega) = \tfrac{1}{2}[S(\omega) + jS_H(\omega)] \tag{2}$$

where $S(\omega)$ is the real spectrum, i.e., the Fourier transform of $R(m)$, and $S_H(\omega)$ is the Hilbert transform of $S(\omega)$.

Due to the analogy between $S^+(\omega)$ in (2) and the analytic signal used in amplitude modulation, a spectral "envelope" $E(\omega)$ [6] can be defined as

$$E(\omega) = |S^+(\omega)|. \tag{3}$$

Due to the large dynamic range of speech spectra, the envelope $E(\omega)$ strongly enhances the highest power frequency bands with respect to $S(\omega)$ [5]. Consequently, the noise components lying outside the enhanced frequency bands are largely attenuated in $E(\omega)$ with respect to $S(\omega)$, and thus, $E(\omega)$ is more robust to broadband noise than $S(\omega)$. On the other hand, as it is well known, the OSA sequence $R^+(m)$ and the signal $x(n)$ have the same poles [7].

Those two properties, i.e., robustness to noise and pole preservation, suggest that AR parameters of the speech signal can be more reliably estimated from the OSA sequence $R^+(m)$ than directly from the signal $x(n)$ when $x(n)$ is corrupted by broadband noise. Thus, as the conventional LPC technique assumes an all-pole model for the speech spectrum $S(\omega)$, we may apply linear prediction to the OSA sequence, assuming an all-pole model for its "spectrum" $E^2(\omega)$. This is the basis of the one-sided autocorrelation linear predictive coding (OSALPC) parameterization technique [5].

A straightforward algorithm is proposed in [5] that calculates the OSALPC cepstral coefficients. It consists of applying the (windowed) autocorrelation method of linear prediction to an estimation of the OSA sequence:

a) First, from the speech frame of length N, the autocorrelation lags until $M = N/2$ are computed (this value of M was empirically optimized to consider the well-known tradeoff between variance and frequency resolution of the spectral estimate [8]).
b) Second, the Hamming window from $m = 0$ to M is applied on such estimated OSA sequence.
c) Third, if p is the prediction order, the first $p+1$ autocorrelation values of that OSA sequence are computed from $m = 0$ to p, using the conventional biased estimator, i.e., the one that is commonly employed in speech processing.
d) Then, these values are used as entries to the Levinson–Durbin algorithm to estimate the AR parameters ak, $k = 1, \cdots, p$.

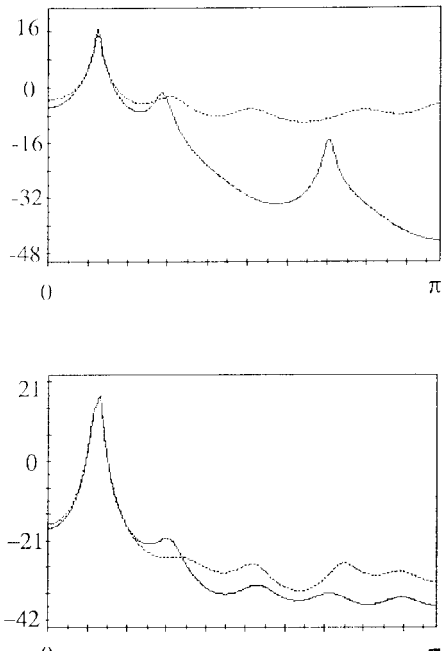

Fig. 1. Robustness of the OSALPC representation to additive white noise: (a) LPC spectrum and (b) OSALPC squared envelope of a voiced speech frame in noise free conditions (solid line) and SNR equal to 0 dB (dotted line).

e) Finally, the cepstral coefficients corresponding to the model are recurrently computed from those AR parameters.

The robustness of OSALPC to additive white noise is illustrated in Fig. 1. As can be seen in this figure, the OSALPC squared envelope shows a prominent first formant, and its whole curve is more robust to additive white noise than that of the LPC spectrum. In this case, the conventional biased autocorrelation estimator was used to compute the OSA sequence from the signal.

Fig. 1 also shows that spurious peaks may appear in the OSALPC square envelope. They are probably due to the fact that the OSALPC technique performs only a partial deconvolution of the speech signal [9]. In spite of that, OSALPC shows a better speech recognition performance than conventional LPC in severe conditions of additive white noise, as will be seen in the next section.

The OSALPC technique is closely related to the short-time modified coherence (SMC) representation proposed by Mansour and Juang in [3]. SMC is also based on AR modeling in the autocorrelation domain. However, whereas in the OSALPC technique, the entries to the Levinson–Durbin algorithm (first p values of the autocorrelation of the OSA sequence) are calculated from the OSA sequence using the conventional biased autocorrelation estimator, in the SMC representation, they are computed using a square root spectral shaper. In fact, in terms of the above formulation, that difference lies in assuming in the SMC technique an all-pole spectral model for the envelope $E(\omega)$ instead of $E^2(\omega)$. Furthermore, $R^+(0)$ is set to 0 in the case of additive white noise because it is severely corrupted by noise.

On the other hand, the name of the SMC representation derives from the usage of a particular estimator, which is referred to as coherence in [3], to compute the OSA sequence from the signal. This estimator is a more homogeneous measure than the conventional biased autocorrelation estimator in the sense that every estimated value is computed using the same number of signal samples, whereas in the conventional estimator, the number of signal samples employed to estimate $R(m)$ decreases along the index m. That property does not have much relevance in the estimation of the autocorrelation entries to the Levinson–Durbin algorithm since only the first $p + 1$

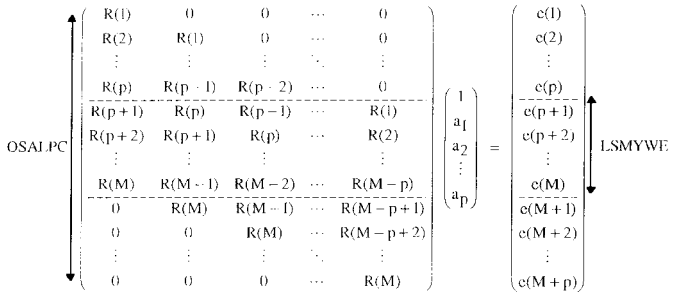

Fig. 2. Matrix formulation for OSALPC and LSMYE methods.

values are considered, and usually, $p \ll N$. However, it may be important in the estimation of the OSA sequence from the speech signal since the OSA length considered in both OSALPC and SMC techniques is $M = N/2$ and not negligible with respect to N.

The OSALPC technique can also be easily related to the overdetermined set of Yule–Walker equations proposed by Cadzow in [4] to seek ARMA models of time series. As an $AR(p)$ process contaminated by additive white noise becomes an ARMA(p, p) process, Cadzow's method can be used to estimate the parameters of this noisy AR process simply by setting the same AR and MA orders in the so-called least squares modified Yule–Walker equations (LSMYWE's) [8].

The relationship between the OSALPC and LSMYWE techniques is illustrated by the matrix equation in Fig. 2, where M denotes the highest autocorrelation lag used, and $e(m)$ is the error to be minimized. The minimization of the norm of the full error vector $\{e(m)\}_{m=1,\cdots,M+p}$ with respect to the AR parameters a_k is equivalent to the application of the (windowed) autocorrelation method of linear prediction to the sequence $R(m)$, $m = 1, \cdots, M$, i.e., the OSALPC technique. On the other hand, the LSMYWE technique minimizes the norm of the subvector $\{e(m)\}_{m=p+1,\cdots,M}$, and therefore, it amounts to applying the (unwindowed) covariance method of linear prediction on the same range of autocorrelation lags. When $M = 2p$, LSMYWE are the modified Yule–Walker equations [8] for an ARMA(p, p) process. In both cases, only autocorrelation lags corresponding to the OSA sequence are employed.

In our comparison, we will also consider another version of this (unwindowed) covariance-based approach that will be called least squares Yule–Walker equations (LSYWE's). Whereas in the LSMYWE technique the first predicted autocorrelation value is $R(p+1)$, in the LSYWE technique, the prediction begins at $R(1)$. Both LSMYWE and LSYWE methods and their relationship to OSALPC are graphically described in Fig. 3. As it is shown, the only difference between the various techniques is the range of autocorrelation lags considered in the minimization of the error. It is worth noting that LSYWE considers some negative autocorrelation lags that do not belong to the OSA sequence. In particular, if M is equal to p, LSYWE are the conventional Yule–Walker equations.

As will be seen in the next section, in spite of the similarity between all these techniques, the OSALPC representation outperforms the LSYWE, LSMYWE, and SMC techniques in speech recognition in severe noisy conditions. On the other hand, as far as the computational complexity of the algorithms is concerned, OSALPC and SMC techniques are much more efficient than LSYWE and LSMYWE techniques because they use the Levinson–Durbin algorithm.

Finally, it is worth noting that the OSALPC technique may be included in the field of higher order spectral estimation due to the fact that the squared envelope $E^2(\omega)$ is the Fourier transform of the autocorrelation of the OSA sequence, which is a particular fourth-order moment of the signal.

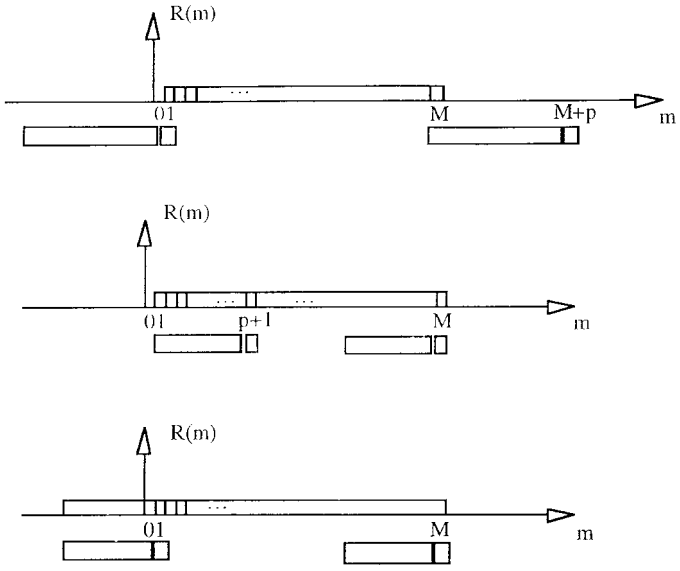

Fig. 3. Interpretation of the (a) OSALPC, (b) LSMYWE, and (c) LSYWE approaches as application of the autocorrelation or covariance methods of linear prediction to an autocorrelation sequence in different lag ranges.

III. Speech Recognition Experiments

This section reports the application of all the above parameterization techniques to recognize isolated words in a multispeaker task with a discrete HMM-based system in order to compare their performance and to gain some insight into the merit of the OSALPC representation in the presence of additive white noise

A. Speech Database and Recognition System

The database used in our experiments consists of 10 repetitions of the Catalan digits uttered by seven male and three female speakers (1000 words) and recorded in a quiet room. First, the system was trained with half of the database and tested with the other half. Then, the roles of both halves were changed, and the reported results were obtained by averaging those two results.

The analog speech signal was first bandpass filtered to 100–3400 Hz by an antialiasing filter, sampled at 8 kHz and, 12 bits quantized. The digitized clean speech was manually endpointed to determine the boundaries of each word. The endpoints obtained in this way were used in all our experiments, including those in which noise was added to the signal. Clean speech was used for training in all the experiments. Noisy speech was simulated by adding zero mean white Gaussian noise to the clean signal so that the SNR of the resulting signal becomes ∞ (clean), 20, 10, and 0 dB. No preemphasis was performed.

In the parameterization stage of the recognition system, the signal was divided into frames of 30 ms at a rate of 15 ms, and each frame was characterized by its cepstral parameters obtained either by the conventional LPC method or by any of the techniques presented in the last section. Before entering the recognition stage, the cepstral parameters were vector quantized using both a codebook of 64 codewords and the Euclidean distance measure between liftered cepstral vectors. Each digit was characterized by a left-to-right discrete hidden Markov model of 10 states without skips. Training and testing were performed using Baum–Welch and Viterbi algorithms, respectively.

B. Recognition Results

First of all, we carried out some experiments with the above described speech recognition system to optimize the model order and

TABLE I
RECOGNITION RATES OF THE CONVENTIONAL LPC TECHNIQUE FOR SEVERAL PREDICTION ORDER VALUES AND CEPSTRAL LIFTERS

ORDER	LIFTERING / SNR(dB)	CLEAN	20	10	0
8	BANDPASS	99.8	92.8	56.8	27.0
	ISD	99.9	97.7	80.0	37.7
	SLOPE	99.7	95.7	72.3	34.1
12	BANDPASS	99.7	96.2	73.7	29.0
	ISD	99.7	97.8	84.0	41.8
	SLOPE	99.8	98.9	89.5	54.2
16	BANDPASS	100	94.0	60.2	19.6
	ISD	99.9	97.7	73.5	32.3
	SLOPE	99.8	93.2	70.7	41.2

TABLE II
RECOGNITION RATES OF THE CONVENTIONAL LPC, LSMYWE, AND LSYWE TECHNIQUES FOR $p = 12$ AND THE SLOPE LIFTER

PARAM. / SNR(dB)	CLEAN	20	10	0
LPC	99.8	98.9	89.5	54.2
LSMYWE	99.5	97.7	81.3	43.1
LSYWE	99.9	95.9	66.9	31.7

TABLE III
RECOGNITION RATES OF THE CONVENTIONAL LPC, SMC, AND OSALPC TECHNIQUES FOR $p = 12$ AND THE SLOPE LIFTER

PARAM. / SNR(dB)	CLEAN	20	10	0
LPC	99.8	98.9	89.5	54.2
SMC	99.0	97.0	89.2	67.5
OSALPC I	98.6	97.7	94.9	79.0
OSALPC-II	99.4	98.4	94.7	72.2

the type of cepstral lifter in the conventional LPC technique. In Table I, the recognition results for LPC model orders $p = 8$, 12, and 16 and for the bandpass [10], inverse of standard deviation [11] (ISD), and slope [12] lifters are presented. The recognition results show that neither the model order nor the type of cepstral lifter are relevant for our task in noise-free conditions. However, in the presence of noise, the recognition results are very sensitive to both factors.

It is also clear from Table I that the nonsymmetrical lifters—slope and ISD—outperform the bandpass lifter for every model order. This may be due to the fact that in the presence of white noise, the lower order cepstral coefficients are more affected than the higher order ones in the truncated cepstral vector.

The best results for severe noisy conditions—10 and 0 dB of SNR—are obtained using slope lifter and prediction order p equal to 12. The convenience of this relatively high order comes from the fact that the sensitivity of the autocorrelation sequence to additive white noise tends to decrease along the lag index. Model orders that are too high, however, yield poor recognition results since the spectral estimate shows spurious peaks. Actually, recognition rates were calculated using the slope lifter for a large range of values of the model order, and the best results were those obtained for $p = 12$.

In Table II, the recognition rates of conventional LPC, LSMYWE, and LSYWE approaches are presented, using $M = N/2$ and both optimum model order and lifter obtained for the conventional LPC technique, i.e., $p = 12$ and the slope lifter. Obviously, these are not the optimum conditions for each parameterization technique, but the results can help to compare their performance. As can be seen from Table II, the conventional LPC technique outperforms noticeably the other approaches. However, the excellent performance of the LSYWE approach in noise-free conditions is worth noting.

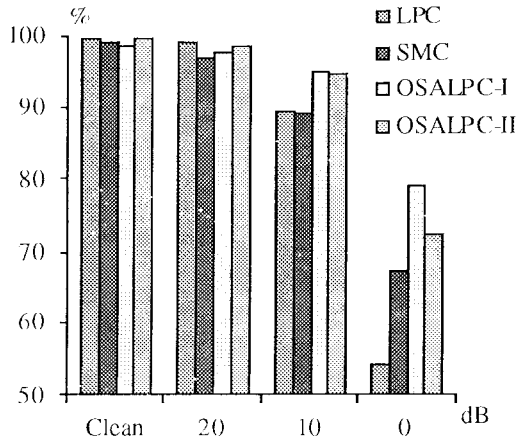

Fig. 4. Comparison of recognition rates of the LPC, SMC, OSALPC-I and OSALPC-II techniques.

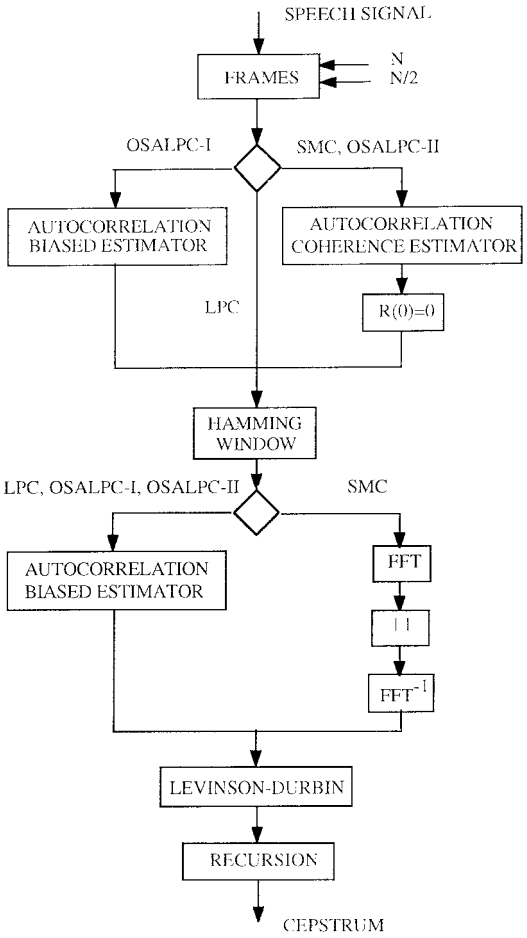

Fig. 5. Block diagram for the calculation of the LPC, SMC, OSALPC-I and OSALPC-II cepstra.

In Table III and Fig. 4, the recognition rates corresponding to the conventional LPC technique, the SMC representation, and the novel OSALPC approach are presented, where we also use $M = N/2$, $p = 12$, and the slope lifter. The two versions OSALPC-I and OSALPC-II of the OSALPC approach correspond to the OSA estimators to which we referred in Section II: OSALPC-I uses the conventional biased autocorrelation estimator, and OSALPC-II like SMC uses the coherence estimator (and sets $R(0)$ to 0). Fig. 5 shows

TABLE IV
RECOGNITION RATES FOR THE OSALPC-II TECHNIQUE FOR SEVERAL PREDICTION ORDER VALUES AND CEPSTRAL LIFTERS

ORDER	LIFTERING / SNR(dB)	CLEAN	20	10	0
8	BANDPASS	97.3	95.5	82.6	44.2
	ISD	97.0	96.4	86.4	52.5
	SLOPE	97.6	97.0	92.5	76.0
12	BANDPASS	98.8	97.2	94.1	71.1
	ISD	98.8	98.3	93.3	68.4
	SLOPE	99.4	98.4	94.7	72.2
16	BANDPASS	99.3	98.7	94.4	76.8
	ISD	99.1	98.1	92.4	72.7
	SLOPE	99.1	98.1	90.7	68.3

a block diagram for the calculation of the LPC, SMC, OSALPC-I, and OSALPC-II cepstra that permits comparison of their respective algorithms.

The OSALPC and SMC representations clearly outdo the conventional LPC technique in severe noisy conditions: OSALPC-I and OSALPC-II rates are better than LPC ones at 10 and 0 dB, and SMC outperforms LPC at 0 dB. Moreover, OSALPC-I and OSALPC-II representations outperform the SMC technique in all noisy conditions. For the OSALPC representation, the use of the conventional biased autocorrelation estimator for computing the OSA sequence (version OSALPC-I) is convenient in severe noisy conditions, i.e., for an SNR of 10 or 0 dB.

However, in noise-free conditions, there is a loss of recognition performance in the OSALPC and SMC approaches with respect to the conventional LPC technique due to the imperfect deconvolution of the speech signal performed by those techniques. This effect seems to be minimized by using the coherence estimator to compute the OSA sequence, as in the case of OSALPC-II and SMC.

Finally, Table IV shows the recognition rates corresponding to OSALPC-II for the same model orders and cepstral lifters as in Table I. It can be noticed that the new technique is less sensitive to changes in both the model order and the type of cepstral lifter than the conventional LPC approach, provided that the model order is not too low.

IV. CONCLUSIONS

In this correspondence, several LPC-based techniques that work in the autocorrelation domain are presented and compared in noisy speech recognition. The OSALPC technique, which is based on the application of the (windowed) autocorrelation method of linear prediction to the one-sided autocorrelation sequence, yields the best results among all the compared LPC-based techniques in severe noisy conditions.

REFERENCES

[1] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67–72, 1975.
[2] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 357–366, 1980.
[3] D. Mansour and B. H. Juang, "The short-time modified coherence representation and its application for noisy speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 795–804, 1989.
[4] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, 1982.
[5] J. Hernando and C. Nadeu, "Speech recognition in noisy car environment based on OSALPC representation and robust similarity measuring techniques," in *Proc. ICASSP'94*, Adelaide, Apr. 1994, pp. 69–72.
[6] M. A. Lagunas and M. Amengual, "Non-linear spectral estimation," in *Proc. ICASSP'87*, Dallas, Apr. 1987, pp. 2035–2038.
[7] D. P. McGinn and D. H. Johnson, "Reduction of all-pole parameter estimation bias by successive autocorrelation," in *Proc. ICASSP'83*, Boston, Apr. 1983, pp. 1088–1091.
[8] S. L. Marple, Jr., Ed., *Digital Spectral Analysis with Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1987.
[9] C. Nadeu, J. Pascual, and J. Hernando, "Pitch determination using the cepstrum of the one-sided autocorrelation sequence," in *Proc. ICASSP'91*, Toronto, Canada, May 1991, pp. 3677–3680.
[10] B. H. Juang, L. R. Rabiner, and J. G. Wilpon, "On the use of band-pass liftering in speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 947–954, 1987.
[11] Y. Tohkura, "A weighted cepstral distance measure for speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1414–1422, 1987.
[12] B. A. Hanson and H. Wakita, "Spectral slope distance measures with linear prediction analysis for word recognition in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 968–973, 1987.

A Fast Algorithm for Finding the Adaptive Component Weighted Cepstrum for Speaker Recognition

Mihailo S. Zilovic, Ravi P. Ramachandran, and Richard J. Mammone

***Abstract*— In speaker recognition systems, the adaptive component weighted (ACW) cepstrum has been shown to be more robust than the conventional linear predictive (LP) cepstrum. The ACW cepstrum is derived from a pole-zero transfer function whose denominator is the pth-order LP polynomial $A(z)$. The numerator is a $(p-1)$th-order polynomial that is up to now found as follows. The roots of $A(z)$ are computed, and the corresponding residues obtained by a partial fraction expansion of $1/A(z)$ are set to unity. Therefore, the numerator is the sum of all the $(p-1)$th-order cofactors of $A(z)$. In this correspondence, we show that the numerator polynomial is merely the derivative of the denominator polynomial $A(z)$. This greatly speeds up the computation of the numerator polynomial coefficients since it involves a simple scaling of the denominator polynomial coefficients. Root finding is completely eliminated. Since the denominator is guaranteed to be minimum phase and the numerator can be proven to be minimum phase, two separate recursions involving the polynomial coefficients establishes the ACW cepstrum. This new method, which avoids root finding, reduces the computer time significantly and imposes negligible overhead when compared with the approach of finding the LP cepstrum.**

I. INTRODUCTION

Speaker recognition is the task of identifying a speaker by his or her voice [1]. A common problem in realizing robust speaker recognition systems is that a mismatch in training and testing conditions seriously degrades the performance [2]. One of the pursued approaches to

Manuscript received February 3, 1995; revised July 13, 1996. The associate editor coordinating the review of this paper and approving it for publication was Dr. Joseph Campbell.

M. S. Zilovic is with Bell Communications Research, Red Bank, NJ USA.
R. P. Ramachandran and R. J. Mammone are with the CAIP Center, Department of Electrical Engineering, Rutgers University, Piscataway, NJ 08855 USA.

Publisher Item Identifier S 1063-6676(97)00762-1.