A Study of LPC Analysis of Speech in Additive Noise

JOSEPH TIERNEY, MEMBER, IEEE

Abstract—A study of the autocorrelation LPC analysis of speech in additive noise is presented. In the noise-free case it is shown that finite word length implementation of the analysis may produce stable but poor spectral estimates. The beneficial effects of proper preemphasis are reaffirmed in terms of decreased numerical error as well as decreased LPC order needed for a good spectral fit. For the case of noisy input speech the conditions for severe distortion of the spectral estimate are presented. A proper LPC spectral analysis of speech in additive noise is shown to require a higher order fit than currently used, a more precise implementation, and a more accurate parameter quantization for transmission.

I. Introduction

The basic autocorrelation LPC vocoder [1] analysis process has been studied in this paper from the point of view of producing more intelligible synthetic speech in the presence of additive input acoustic noise. The tools used in this study were: 1) sliding analysis LPC outputs representing reflection coefficients estimated from data windows shifted by four sample points for each analysis, 2) synthetic all-pole vowel waveforms to which noise of various spectral shapes and amplitudes was added, and 3) display of the resulting LPC spectral match. Preliminary studies, using these tools, were made of the autocorrelation procedure and Levinson recursion implemented in fixed-point arithmetic for practical analysis-synthesis. Poor spectral estimates were seen to arise in some cases from the numerical noise produced by 16 bit fixed-point procedures. The numerical noise, which is cumulative, shows up more severely on higher order k's so that these parameters can be badly distorted, but still produce stable estimates.

The study of the noise-free case provided considerable insight for the study of the noisy input case. For the case of an additive noise whose spectrum is flat (white), the LPC spectrum fit becomes inaccurate at the upper end of the voice spectrum. As the additive noise spectrum becomes more shaped (e.g., white noise filtered by a complex pole pair), its effect upon the speech LPC spectrum fit becomes more pronounced. In principle, the nonflat noise spectrum is "using up" some of the poles which should be matched to the speech spectrum. In order to provide more poles to fit both noise and speech, a higher order LPC fit must be made. However, using a higher order system is only part of the solution. Since the higher order k's are still needed to model the more complex speech plus noise spectrum, these k's are not close to zero, but may be quite larger, albeit with magnitudes less than 1. In order to reproduce this high-order spectrum at the receiver, there must be more accurate transmission coding of these upper k's than is typically provided in most 2400 bits/s LPC coding schemes. As in the noise-free case, numerical accuracy becomes an issue, and a fixed word length of more than 16 bits or a double precision realization is necessary.

A suitable LPC vocoder for noisy acoustic environments presenting the LPC analyzer with speech corrupted by additive noise should have the following properties:

1) an analysis of sixteenth- to eighteenth-order for a 125 μs sample rate;

2) an implementation which retains accuracy for this high-order model;

3) a method of quantization for transmission which does not distort higher order k's close to ± 1.

In Section II we discuss the noise-free studies of the LPC analysis process using the aforementioned tools. In Section III we report results of the study of noisy inputs. Finally, in Section IV we discuss similar issues in channel vocoders and noise preprocessors.

II. LPC Analysis with Noise-Free Input

In order to study the effect of additive noise upon LPC analysis, we first studied the analysis of clear speech. A quasi-continuous 16 bit fixed-point LPC analysis routine was implemented which produced a set of reflection coefficients for a 160 point data window (20 ms at 125 μs sampling) and then shifted the window by 4 points for another analysis. This analysis consists of a Hamming window operation (single precision), autocorrelation (double precision accumulate and block scaling to single precision), and a Levinson recursion (single precision). Using a tenth-order analysis we obtained plots of continuous reflection coefficients (k's) for various input utterances. Fig. 1(a)–(c) shows the nonpreemphasized utterance "rare" for a male speaker along with the second and sixth reflection coefficients (k_1 and k_5) of the analysis. The coefficient k_1 is well behaved, indicating some movement on the fourth line. The coefficient k_5 shows even more movement and also what appears to be noise especially on the first line. From Markel's data on fixed-point implementation [2], our 16 bit implementation with block floating correlation points and no preemphasis should be producing stable tenth-order LPC estimates. If we compute k_5 using a full floating-point implementation [see Fig. 1(d)], and com-

Manuscript received May 25, 1979; revised October 11, 1979 and February 10, 1980. This work was supported by the U.S. Air Force. The views and conclusions in this paper are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

The author is with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173.

0096-3518/80/0800-0389$00.75 © 1980 IEEE

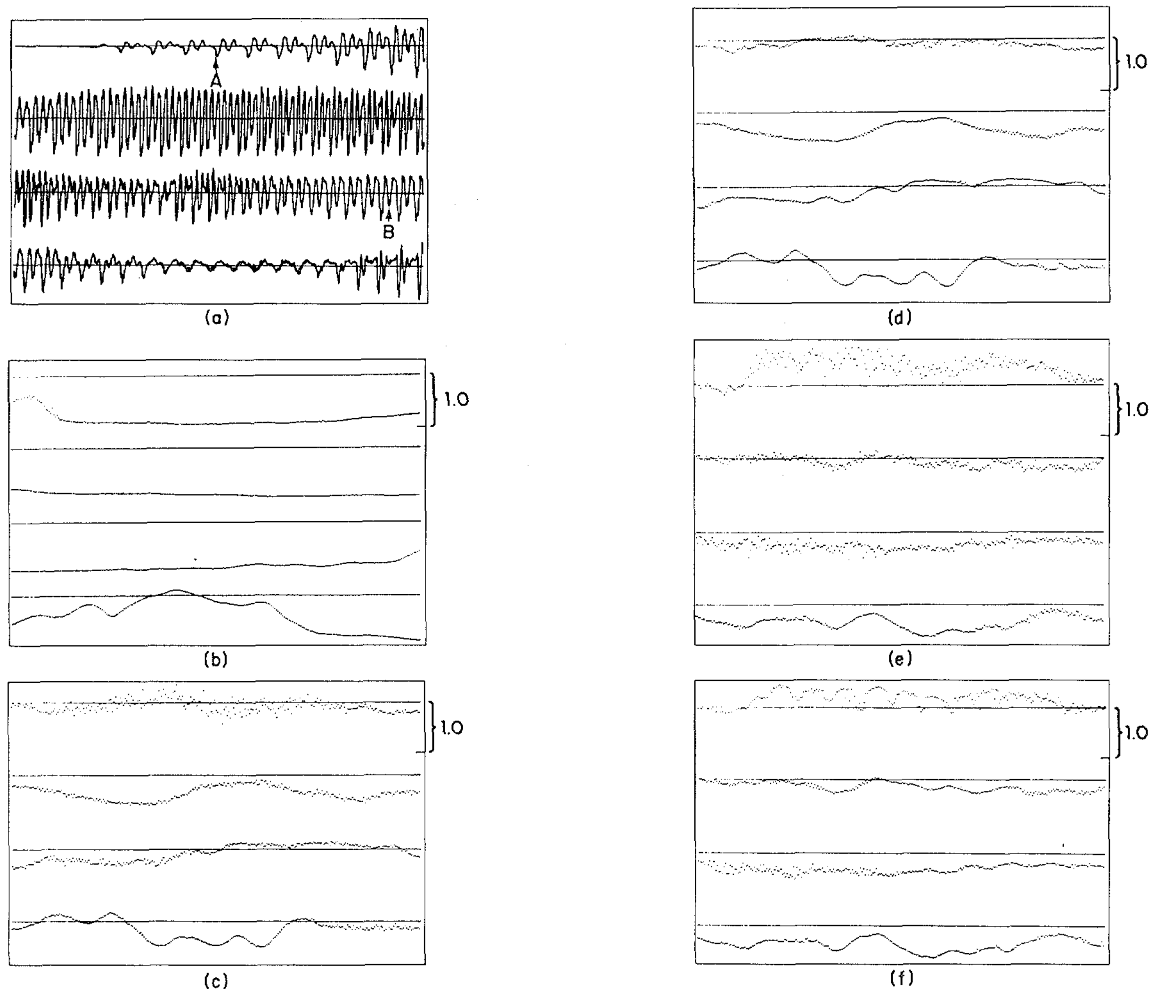


Fig. 1. (a) Unpreemphasized male utterance "rare" (1024 points/line 125 μs sampling). (b) Sliding analysis k_1 (fixed point). (c) Sliding analysis k_5 (fixed point). (d) Sliding analysis k_5 (floating point). (e) Sliding analysis k_9 (fixed point). (f) Sliding analysis k_9 (floating point).

pare Fig. 1(d) and (c), we see the noise introduced by our 16 bit implementation. If we examine the tenth coefficient k_9 for fixed- and floating-point analysis [see Fig. 1(e) and (f)], we again see the noise introduced by the fixed-point system, whereas the k_9 plot for floating-point analysis only shows some pitch period variation. K_9 has been reset to zero several times during the fixed-point analysis [difficult to see in Fig. 1(e)], which indicates that the computation noise has built up enough to drive $|k_9| \geqslant 1$, and the routine has aborted at a lower order fit. Since our analysis is performed by sliding every 4 points (for an overlap of 156 points) instead of every 160 points (for a typical 50 Hz update rate), we are seeing a 40 times denser analysis activity. A normal analysis might show 1 or 2 instabilities at most. It is interesting that we have implemented an LPC analysis in fixed-point that appears adequate from all considerations of stability, but clearly introduces significant noise into the estimates of the k's. Fig. 2 shows the DFT spectra resulting from using successive k's (16 bit) in this noisy region to synthesize the unit sample response of a tenth-order acoustic tube. A corresponding floating-point LPC fit is also shown for comparison. Estimates using the floating-point output are less noisy and more consistent although perceptual differences due to these "noisy" spectra may be less pronounced than they appear to be. Note that these spectrum variations are different in origin from the Q variations reported in the literature [3]-[5] due to window position and width. The variations we observe are caused predominantly by the fixed-point numerical error. For comparison, the less noisy k's from the fourth line fixed-point analysis produce the spectra in Fig. 3. If the analysis window is doubled to 320 points, a predictable smoothing of the k track will occur, but there is no decrease in the finite word length analysis noise. Since the noise is generated in the recursion and not the autocorrelation operation, this is not surprising.

Markel and others [6], [7] suggest the use of preemphasis prefiltering to condition the input speech before analysis. There are several justifications for this operation. From a theoretical point of view, a proper prefilter may remove the effects of glottal waveshape and the radiation characteristic of the lips. This will leave the all-pole vocal tract filter for analysis without wasting the LPC poles on glottal and radiation shaping. From a spectrum point of view, any preliminary flattening of the overall input spectrum before LPC processing allows the LPC analysis to do its own job of spectrum flattening better. Basically, these two statements imply that proper speech preemphasis will reduce the order of an LPC fit needed

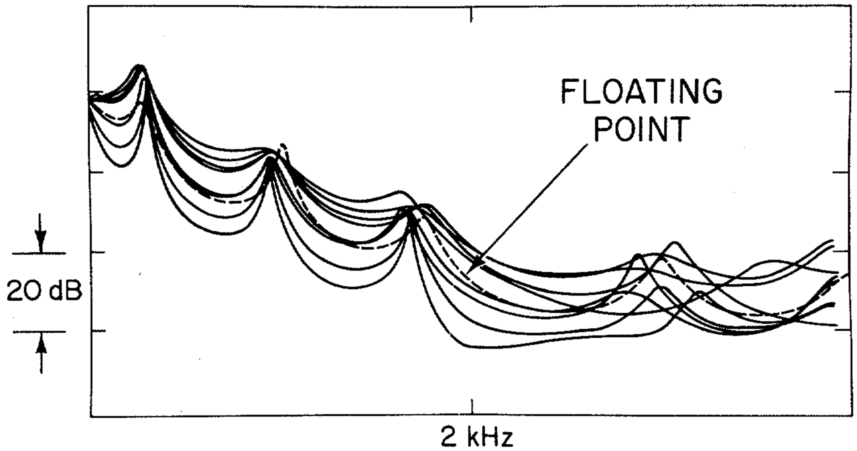


Fig. 2. DFT spectra from noisy fixed-point LPC fits.

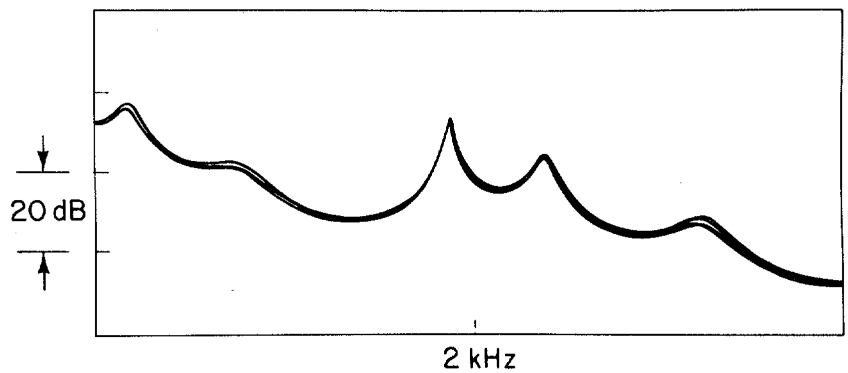


Fig. 3. DFT spectra from less noisy fixed-point LPC fits.

to do an equivalent spectrum match. Finally, from the point of view of a finite word length implementation, the proper preemphasis will reduce numerical error. This can be seen from expressions for the first two reflection coefficients in terms of autocorrelation points

$$k_0 = \frac{R_1}{R_0} \tag{1}$$

$$k_1 = \frac{R_2 - k_0 R_1}{R_0(1 - k_0^2)}. \tag{2}$$

The expression for k_1 can be rewritten, using the value of k_0, as

$$k_1 = \frac{R_0 R_2 - R_1^2}{(R_0 - R_1)(R_0 + R_1)}. \tag{3}$$

From (3) we see that if $R_1 \approx R_0$

$$k_1 = \frac{(R_2 - R_1)}{2(R_0 - R_1)} \tag{4}$$

and since

$$|k_1| \leqslant 1$$

$$|R_2 - R_1| \leqslant 2|R_0 - R_1| \tag{5}$$

so that if $|R_0 - R_1|$ is small then $|R_2 - R_1|$ is small also. For a finite word length implementation (e.g., 16 bits) this implies that both numerator and denominator of (4) are strongly affected by the 16 bit truncation error. This situation leads to a less accurate k_1 than would be the case for R_0 and R_1 not close in value.

We can write the correlation points as a function of the input speech power spectrum as

$$R_k = \frac{1}{\pi} \int_0^{\pi} S(\theta) \cos(K\theta)\, d\theta \tag{6}$$

where $S(\theta)$ is the power spectrum over the interval $(0, \pi)$ (0–4000 Hz). The ratio R_1/R_0 then becomes

$$\frac{R_1}{R_0} = \frac{\int_0^{\pi} S(\theta) \cos\theta\, d\theta}{\int_0^{\pi} S(\theta)\, d\theta}. \tag{7}$$

If the spectrum is concentrated around low frequencies (e.g., a delta function close to zero frequency), the ratio approaches one. If the energy is concentrated around 4000 Hz (π), the ratio approaches minus one. Both of these cases will produce large fixed-point error as a result of the aforementioned mechanism. A flat spectrum will produce a ratio approaching zero leading to small error. Therefore, it follows that the speech spectrum should be shaped to flatten the energy over the widest frequency range since this will make R_1 closer to zero and, hence, move R_1/R_0 away from unity magnitude. This justifies the need for preemphasis from a low error point of view.

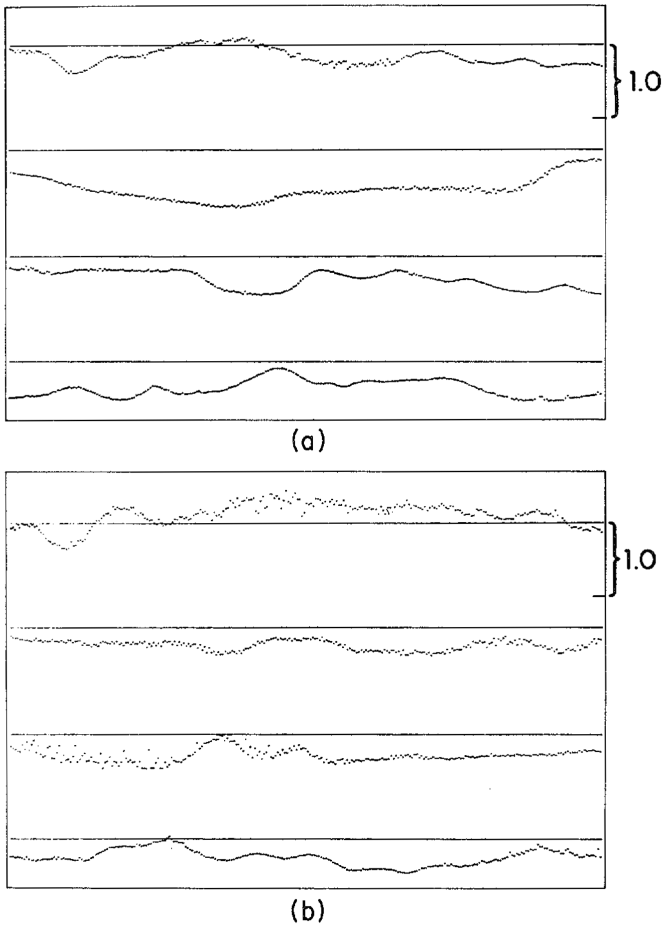


Fig. 4. (a), (b) Reflection coefficients k_5 and k_9 of preemphasized signal.

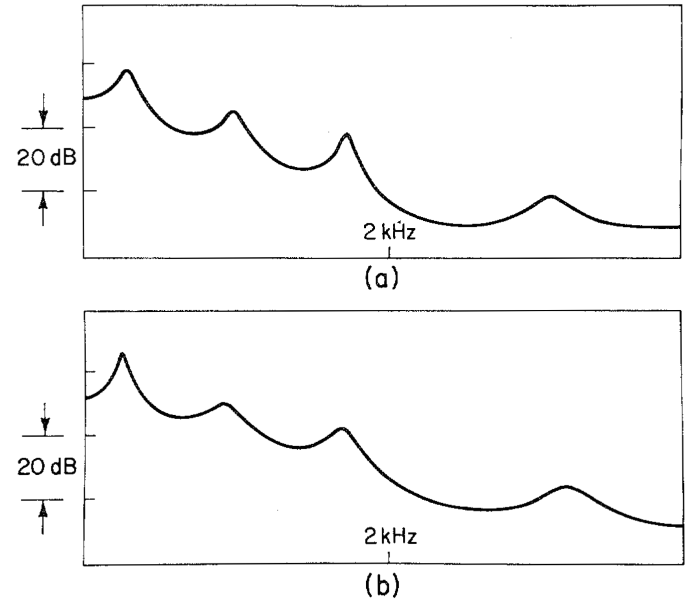


Fig. 5. (a), (b) Spectra of preemphasized speech with $k_9, k_9 = 0$.

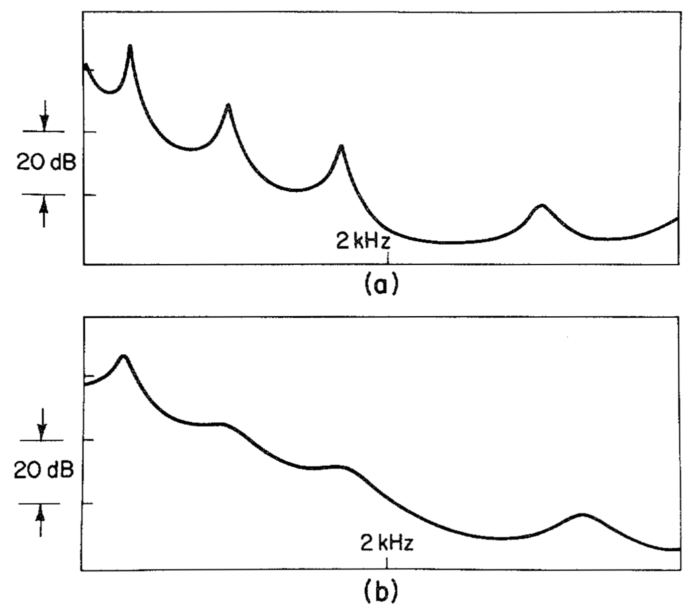


Fig. 6. (a), (b) Spectra of unpreemphasized speech with $k_9, k_9 = 0$.

Fig. 4(a) and (b) indicates the decreased numerical error associated with k_5 and k_9 of a properly preemphasized signal. The speech signal is the same as that of Fig. 1 with a boost of 6 dB/octave starting at 400 Hz (real axis zero at 400 Hz). If we examine k_5 and k_9 for this fixed-point analysis, they appear much less noisy than the unpreemphasized case. To show that this preemphasis has reduced the requirements for a good spectrum match, as well as reducing noise, Figs. 5(a) and 6(a) represent LPC spectra (DFT of acoustic tube impulse response) for the preemphasized and unpreemphasized speech. Figs. 5(b) and 6(b) are the same spectra with k_9 set to zero; notice how this affects the preemphasized spectrum less, indicating that a lower order match is now sufficient. It must also be noted that improper preemphasis may hinder the LPC analysis by the same mechanism. If the input signal spectrum has already been shaped by microphone characteristics (e.g., noise-cancelling microphones) or some spurious linear filter (e.g., a poor telephone line) then the standard analog preemphasis by a real axis zero in the frequency domain (or the digital equivalent) may in fact increase the complexity of the signal spectrum causing a poorer LPC spectrum fit.

To examine more closely the properties of LPC analysis, we have generated several synthetic vowels by using the reflection coefficients from various points of our analysis to set up acoustic tube filters driven by uniform pitch periods of 96, 48, and 32 samples (12, 6, and 4 ms, respectively). One set of coefficients was used from Fig. 1, point A, another from point B. The first is a very noisy set of coefficients, the second much less so. Figs. 7 and 8 are the synthetic waveforms for 6 ms periods, as well as the continuous analysis results of those waveforms (k_5 and k_9). Notice how much noisier the k estimates of the first vowel are compared to the second. This is borne out by Figs. 9 and 10, which are LPC spectra represented by some of the reflection coefficients of Figs. 7 and 8. If we compare k_9 for each of these vowels, we see that k_9 for the noisier vowel is greater (i.e., closer to one) than for the well-behaved vowel. This is true for these synthetic vowels for pitch periods of 4 and 12 ms as well. It appears that the buildup of the noise in the 16 bit implementation will not perturb an estimate whose last few reflection coefficients are small and noncritical, but will be responsible for serious spectrum distortion if these coefficients are close to ± 1 and, therefore, important to the spectrum match [8], [9]. A proper preemphasis will reduce this distortion in two ways. First, because of the numerical noise argument presented earlier, and second, because of the flatter lower order spectrum resulting from proper

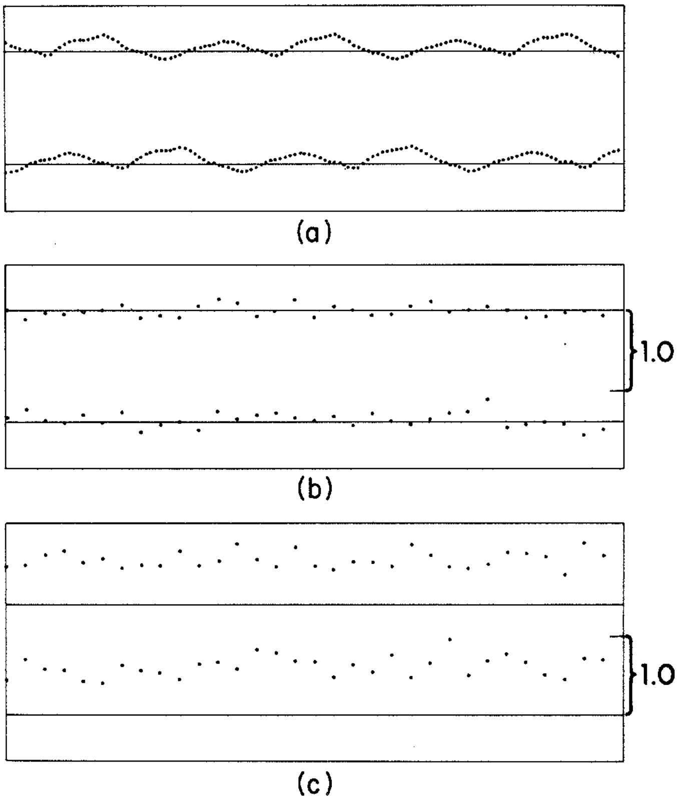


Fig. 7. (a)–(c) Synthetic vowel waveform from k's point A, k_5 and k_9.

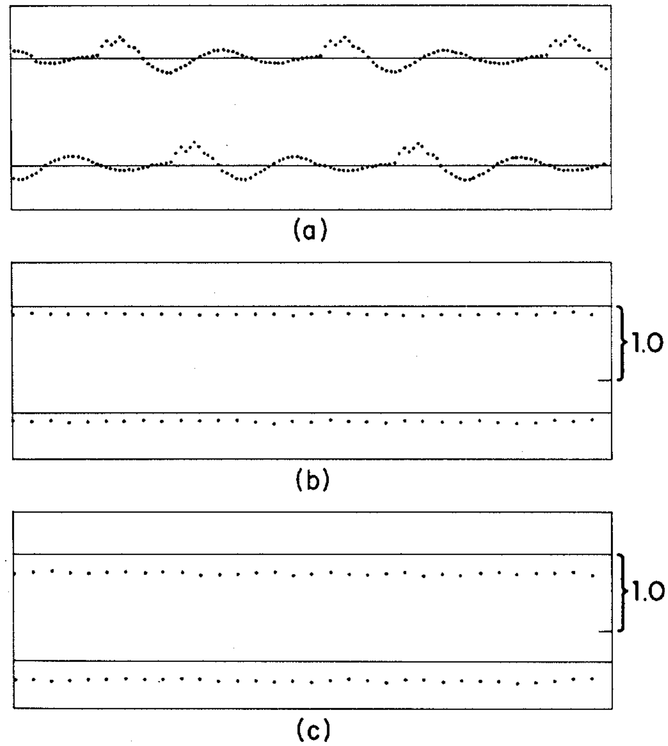


Fig. 8. (a)–(c) Synthetic vowel waveform from k's point B, k_5 and k_9.

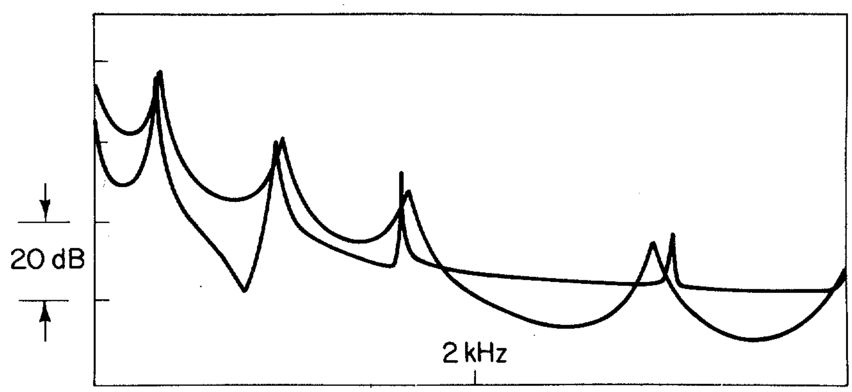


Fig. 9. Spectra from analysis of synthetic vowel, Fig. 7.

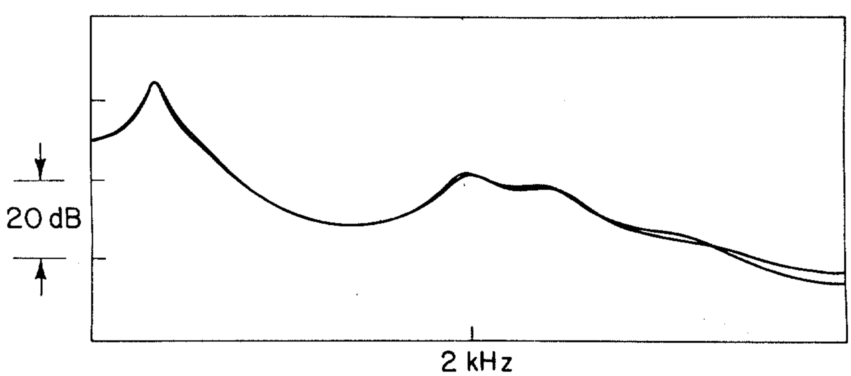


Fig. 10. Spectra from analysis of synthetic vowel, Fig. 8.

preemphasis. A floating-point or double precision implementation can overcome the numerical noise problem, but the upper k's will not be close to zero and will be sensitive to quantization introduced for transmission.

We can conclude that present finite word length autocorrelation LPC systems using 16 bit words and running at 8 kHz or less can adequately represent input speech spectra if proper equalization is used. This will guarantee reasonable numerical accuracy. It will also result in sets of reflection coefficients that are not sensitive to quantization for transmission using only two or three bits for the higher coefficients. If the input speech is not easily equalized because of unknown transmission parameters (e.g., poor audio lines or microphones), or microphone outputs whose frequency response depends on speaker position with respect to the microphone (e.g., noise-cancelling microphones), then other measures are necessary. The general approach to LPC analysis-synthesis for input spectra which may not be equalized so that high-order k's approach zero must include both more accuracy in the analysis implementation (e.g., double precision) and more careful consideration of the quantization distortion allowed for the higher reflection coefficients. In the next section this discussion for noise-free input will be shown to be even more important for the noisy input case.

III. Analysis of Noisy Waveforms

Our synthetic vowel waveforms were modified by adding random noise with a flat spectrum or with a spectrum corresponding to a second-order resonance (noise samples filtered by a second-order filter).

A. Additive White Noise

In the case of additive noise with a flat spectrum the noise sample amplitudes were adjusted to provide a desired signal-to-noise ratio (s/n)

$$\mathrm{s/n} = \frac{R_0}{N\sigma_n^2} \tag{8}$$

where R_0 is the sum of the squared amplitudes of the speech samples, N is the number of samples, and σ_n^2 is the added noise variance.

Performing the LPC analysis on this vowel-noise combination (using the vowel of Fig. 7), with an s/n of 30 dB, we see a broadening of the upper formants in Fig. 11(b) [Fig. 11(a) is the true speech spectrum] and a decrease in the magnitude of k_8, k_9 [see Fig. 11(c) and (d)]. For s/n = 20 dB, we notice the upper formants disappearing completely in Fig. 12 and the upper reflection coefficients approaching zero. This is not surprising since the LPC analysis uses only the input power spectrum (i.e., the correlation function), and this power spectrum has its upper formants flattened out by the noise. To the LPC analysis this spectrum appears as a lower order spectrum than that for the speech alone, and upper reflection coefficients are not needed for a fit. Any preemphasis on this speech plus noise waveform will not reduce this higher formant loss since it will not increase the "local" s/n ratio at these higher formants. In fact, the overall s/n ratio may decrease. The saving feature of this white noise degradation is its "graceful" performance. That is, it slowly decreases the higher formant resolution as the s/n ratio decreases and consequently only slowly decreases intelligibility. There is no nonlinear threshold effect. Ironically, in terms of numerical accuracy considerations, the situation is beneficial since the higher k's are close to zero (i.e., the order of the spectrum has been reduced by obscuring upper formants) and the inaccuracy associated with them does not influence the spectral match.

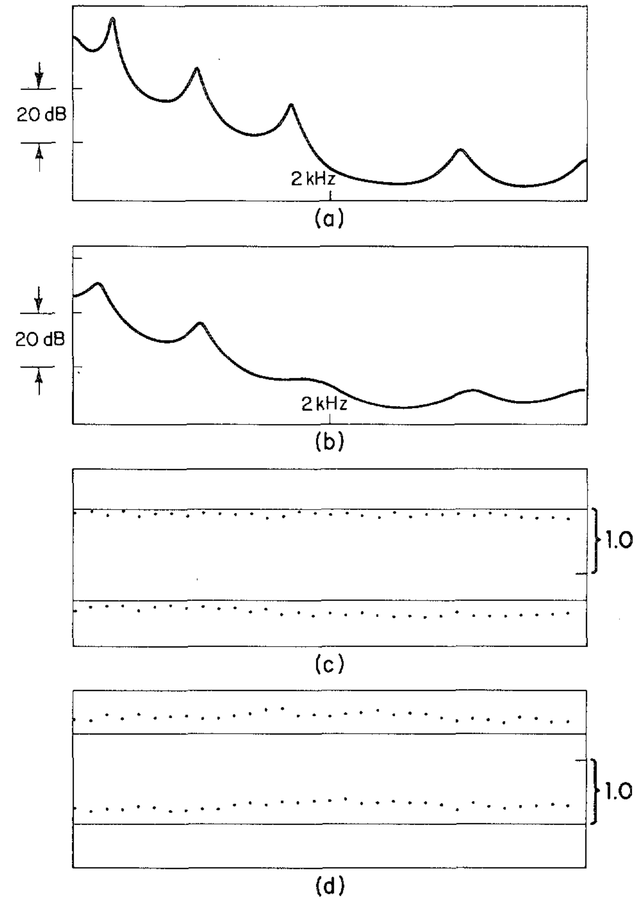


Fig. 11. Analysis of synthetic vowel and noise, s/n = 30 dB. (a) Noise-free spectrum. (b) Noisy spectrum. (c) k_8. (d) k_9.

B. Additive Colored Noise

When the additive noise begins to possess a spectral complexity of its own (i.e., is no longer white) the results of LPC analysis on the composite signal are more interesting. In order to spectrally shape the additive noise, the white noise samples x_n are filtered through a second-order recursion of the form

$$y_n = x_n + 2r \cos bT(y_{n-1}) - r^2 y_{n-2} \tag{9}$$

where b is the resonant frequency, $r = e^{-\alpha T}$, α is one-half the resonant bandwidth, and T is the sampling interval. The filtered noise samples y_n are added to the synthetic vowel waveform to form a speech waveform corrupted by additive colored noise. The s/n ratio is still given by (8) except that σ_n^2 is now the variance of a tuned noise and given by

$$\sigma_n^2 = [\sigma_0^2] \frac{1+r^2}{1-r^2} \frac{1}{1 - 2r^2 \cos 2bT + r^4} \tag{10}$$

with r and b as above and σ_0^2 the input white noise variance.

When the added noise is peaked at 600 Hz with a bandwidth of 200 Hz, the spectral perturbation is very low even for an s/n of 20 dB. Fig. 13(a) shows a sample of the noise spectrum alone, and Fig. 13(b) the LPC spectrum of the signal plus noise. This is not a surprising result if we recognize that the tuned noise is shaped similarly to the spectrum of the noise-free vowel. As a result the added noise spectrum is distributed so as to be largest where the speech spectrum is largest, so that

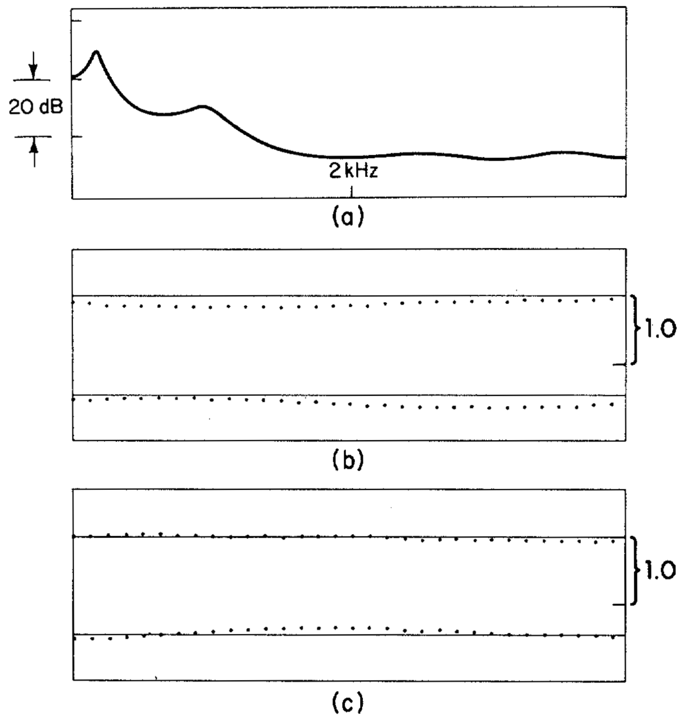


Fig. 12. Analysis of synthetic vowel and noise, s/n = 20 dB. (a) Noisy spectrum estimate. (b) k_8. (c) k_9.

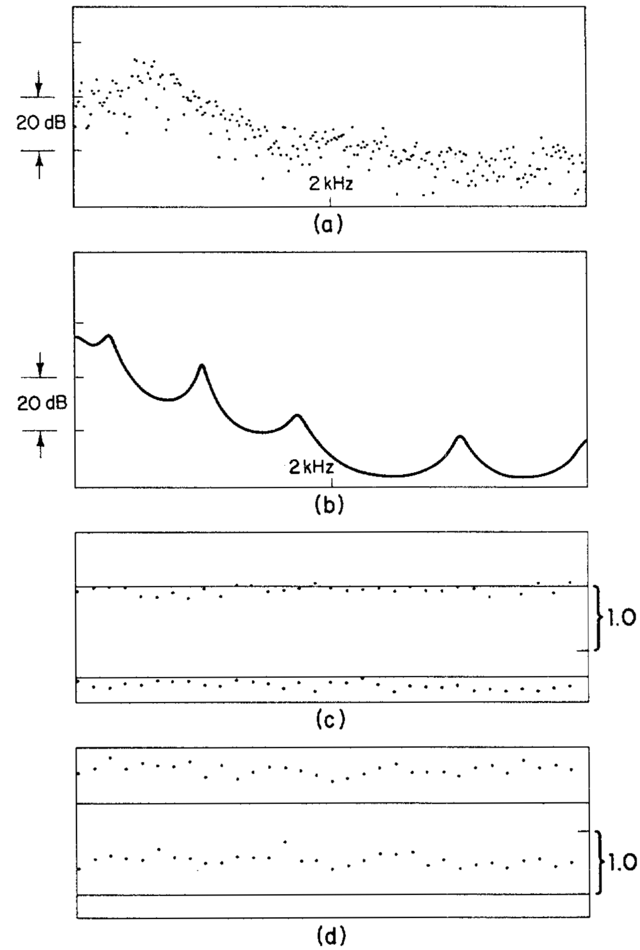


Fig. 13. Analysis of synthetic vowel plus tuned noise. (a) Noise spectrum. (b) LPC analysis spectrum. (c) k_8. (d) k_9.

s/n ratios in local regions are not low. If we inspect the higher reflection coefficients (e.g., k_8 and k_9) in Fig. 13 as well as the LPC spectrum match, we see how little this noise has perturbed the LPC analysis process.

If we tune the noise to a higher frequency resonance (2500 Hz, 200 Hz bw) so as to perturb the lower energy portions of the synthetic vowel, we see a much more destructive result. Fig. 14 shows the noise shape and the LPC fit for s/n = 30 dB. Even for this s/n (toll quality), we have completely lost both third and fourth formants as shown in Fig. 14(b). The LPC analysis has modeled the higher amplitude noise peak at 2500 Hz rather than the upper formants. The noise peak has dominated the third and fourth formant peaks, and the tenth-order LPC analysis does not have enough poles to model the noise peak and the formant peaks simultaneously. Preemphasizing the input spectrum before analysis does not improve the s/n ratio around important regions, but it does serve the purpose in this case of introducing an additional degree of freedom by reducing the complexity of the input spectrum. The resulting analysis of the preemphasized input does indeed show the reappearance of third and fourth formants [see Fig. 14(c)]. This difference between the two fits points to the need for a higher than tenth-order LPC analysis to model the input of synthetic vowel and tuned noise. Fig. 15(a) and (b) shows the enhanced fit possible when a higher order LPC analysis is performed without and with preemphasis, twelfth-order in these two cases. Even the twelfth-order system without preemphasis is not sufficient. The full twelfth-order and the extra preemphasis processing is needed to provide a reasonable spectrum analysis. Inspection of the upper k's in these two cases shows k's significantly bigger than zero contributing to the spectral fit. Therefore the colored noise spectrum is "using up" some of the poles which were needed to match the speech spectrum. Obviously, if the noise is of sufficient complexity and energy, many of the poles used for the speech match will be needed to model the noise spectrum and the resulting speech analysis will be poor. Using a higher order LPC analysis will not eliminate the noise, but it will allow modeling of both speech and noise spectral properties with sufficient accuracy so that high-energy noise peaks requiring several LPC poles for a good fit will not swamp out nearby formants as seen in Fig. 14. At the synthesizer output the higher order modeled speech and noise output will still be influenced by the noise, but will reproduce in a quasi-linear sense the entire input spectrum rather than reproducing a lower order speech plus noise spectrum at the expense of the input speech.

Depending on the complexity and amplitude of the additive noise spectrum, an LPC analysis of sixteenth- or eighteenth-order would be appropriate at an 8 kHz sampling rate (e.g., three simple noise resonances would require 6–8 poles). This higher order analysis (i.e., greater than tenth-order) leads to the need for higher numerical accuracy. As we have pointed out earlier, our tenth-order, 16 bit, fixed-point implementation of the Levinson recursion along with double precision accumulation and block scaling of the autocorrelation points is very close to being marginal. For a sixteenth- or eighteenth-

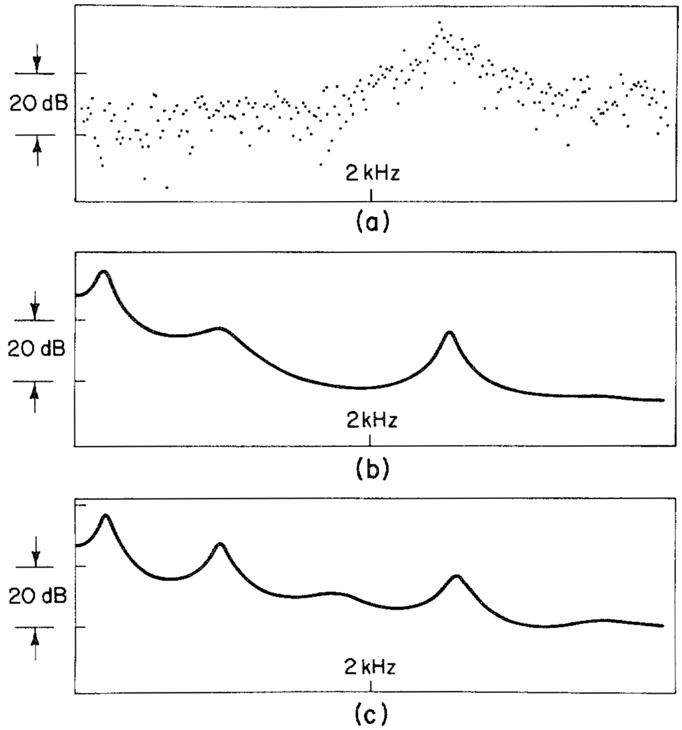


Fig. 14. (a) Noise tuned to 2500 Hz. (b) Analysis of synthetic vowel and noise s/n = 30 dB. (c) Analysis of preemphasized vowel plus noise.

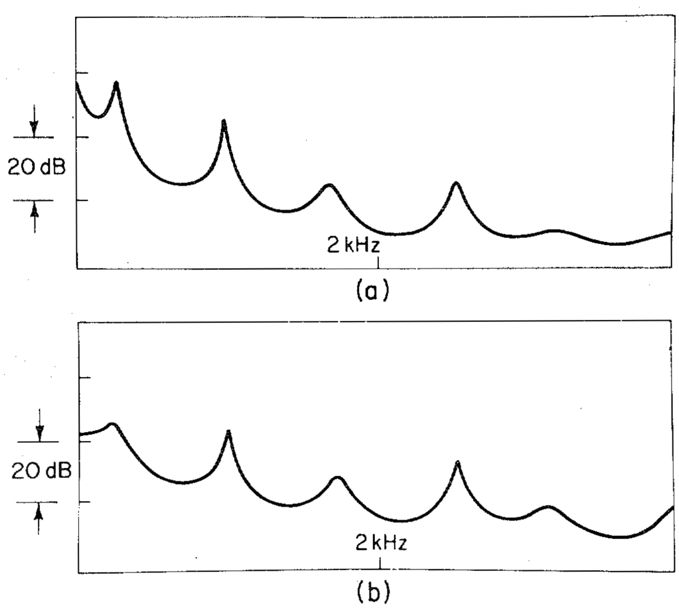


Fig. 15. Twelfth-order LPC analysis spectra of Fig. 14 signal. (a) Without preemphasis. (b) With preemphasis.

order analysis implementation, the correlation point computation could be realized in double precision fixed point with little loss of accuracy (i.e., 32 bit word). The Levinson recursion could also be implemented adequately in double precision, implying double precision multiplies and divides.

Given these more accurate sixteenth- or eighteenth-order analysis spectra characterized by 16 to 18 reflection coefficients, we still face the problem of quantization for transmission. Many current analysis-synthesis LPC systems used at 2400–4800 bits/s use a tenth-order fit and quantize the ten reflection coefficients in fewer bits going from k_0 to k_9. This is based on histograms of the reflection coefficients derived from a large set of input utterances. These histograms and the resultant quantization assignments for tenth-order systems are adequate for noise-free input, reasonably preemphasized. However, if the input spectrum is not so well behaved or has a complicated noise spectrum added to it, the upper k's become much larger in variance. The practical approach to implementation of an LPC system suitable in an acoustic noise environment would include both a higher order analysis (sixteenth- to eighteenth-order) and a more accurate quantization of the higher k's. At present this would seem to indicate a higher data rate than 2400 bits/s. However, models of the input spectrum with this many coefficients may have new properties which allow for advantageous coding, given a suitable transformation of the reflection coefficients. It is worth noting that a channel vocoder of 16 or 18 channels can be quantized to 2400 bits/s and lower. The combination of a low-order model (e.g., tenth-order), a noisy environment, and crude quantization of upper k's can only produce low intelligibility synthetic speech.

IV. Comments on Channel Vocoders and Noise Preprocessors

Raising the order of an LPC analysis in order to accommodate a spectral fit to both the speech signal and additive noise is certainly not as effective as eliminating the input noise before or during the spectrum analysis process. On the other hand, channel vocoders with 18 or 16 filter subchannels seem to provide a more intelligible synthetic output for various kinds of acoustic input noise than do LPC vocoders of tenth-order. At best the channel vocoder is doing a good job of reproducing the input signal-plus-noise spectrum (albeit imposing a harmonic nature on the noise in the case of voiced input). For a vocoder filter bank analysis using filters with sharp skirts (little overlap) noise sharply peaked in one spectral region will not influence more than the spectrum associated with the channel filter for that region. This is different from the LPC analysis which may lose the speech character completely if the noise is of sufficient complexity and amplitude. Of course, any coding and quantization which reduces the quasi-orthogonal nature of the channel analysis outputs, such as a delta coding from channel to channel, will spread the effects of a noise perturbation over a wider spectral region.

Some sparse intelligibility data for a Lincoln Laboratory LPC vocoder using a tenth-order fit at 2400 bits/s and a twelfth-order fit at 4800 bits/s [10] supports our argument for higher order LPC analysis of noisy signals (or at least prompts us to pursue this subject further). For noise-free input speech the vocoder intelligibility increased about three points from 2400 to 4800 bits/s. The difference in the two LPC vocoder implementations was an increase in order from tenth to twelfth and more accurate encoding of reflection coefficients. When the input speech was noisy, intelligibility increases of 7–10 points were reported between 2400 and 4800 bits/s. The increased analysis order (tenth to twelfth) and the extra encoding levels for high-order k's seem to contribute to higher intelligibility with noisy input, as we would expect.

A preprocessing noise-reducing device preceding an LPC vocoder to improve intelligibility has been proposed as a practical solution for improved LPC vocoder robustness [11], [12]. Several approaches to the preprocessing problem have been described. It is important to consider the effect upon the LPC analyzer input spectrum of such a preprocessor. If the noise reduction process increases the complexity of the input signal spectrum, the LPC analysis may require an increase in order, accuracy, and quantization levels. In this case the total solution will be a noise preprocessor followed by a higher order LPC vocoder than the present standard, not unlike our recommendations for LPC analysis of noisy input.

V. Conclusions

In this report we have examined the autocorrelation LPC analysis process spectral fit. We have shown that finite word length (16 bit) realization of the LPC analysis process yields stable but noisy results that can lead to poor spectral estimates of input speech with no additive noise. In particular, input speech signals whose spectra require a fit using upper reflection coefficients (e.g., k_8, k_9) with magnitudes substantially greater than zero are particularly susceptible to this numerical noise. Correct preemphasis reduces both the numerical error and the order of the fit required (i.e., upper k's approach zero). Incorrect preemphasis will increase this effect. For input speech signals whose overall spectrum equalization is questionable, due to time varying microphone or transmission characteristics, a more accurate numerical implementation and a more accurate quantization of upper k's are necessary to retain intelligibility.

The LPC analysis of speech with additive noise is shown to produce adequate spectral modeling if the order of the LPC all-pole filter is high enough to model both speech and noise spectral features. If the LPC order remains at ten and a complex noise signal of sufficient amplitude is present, speech formants may be completely missed in order to model noise peaks. Again the higher order fit must be accompanied by more accurate quantization of upper reflection coefficients.

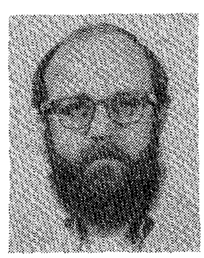
Finally, it is suggested that some of the present approaches towards noise stripping preprocessors presenting input to regular LPC vocoders may be producing input spectra requiring a higher order fit than for the original speech.

Acknowledgment

The basic data handling software for this study was provided by D. Christopher. E. Singer supplied a very useful and flexible spectrum analysis procedure. Discussions with R. McAulay throughout this study have given the author much insight about LPC modeling of speech and noise.

References

[1] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* Berlin, Germany: Springer, 1976, sec. 3.3.
[2] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* Berlin, Germany: Springer, 1976, sec. 9.4.
[3] L. R. Rabiner, B. S. Atal, and M. R. Sambur, "LPC prediction error–Analysis of its variation with the position of the analysis frame," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, p. 441, Oct. 1977.
[4] D. G. Nichol and R. E. Bogner, "Quasi-periodic instability in a linear prediction analysis of voiced speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 210–216, June 1978.
[5] Y. Tohkura, F. Itakura, and S. Hashimoto, "Spectral smoothing techniques in PARCOR speech analysis-synthesis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 587–596, Dec. 1978.
[6] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* Berlin, Germany: Springer, 1976, sec. 6.5.6.
[7] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, p. 561–580, Apr. 1975.
[8] J. D. Markel and A. H. Gray, *Linear Prediction of Speech.* Berlin, Germany: Springer, 1976, sec. 10.2.2.
[9] 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, June 1975.
[10] E. M. Hofstetter *et al.*, "Vocoder implementations on the Lincoln digital voice terminal," in *EASCON '75 Rec.*, 1975, pp. 32A–32J.
[11] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision maximum-likelihood noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 137–145, Apr. 1980.
[12] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113–120, Apr. 1979.



Joseph Tierney (S'54–M'57) was born in New York, NY, on January 14, 1934. He received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1956.

He was a co-op student at M.I.T. at the Bell Telephone Laboratories, and spent the year 1955–1956 as a Research Assistant at the M.I.T. Research Laboratory of Electronics. Since 1961 he has been with the M.I.T. Lincoln Laboratory, Lexington, MA, working in the areas of speech compression and satellite communications at both a systems and circuits level. For the past several years he has been involved in the design of digital signal processing devices.