X. SPEECH COMMUNICATION*

Academic and Research Staff

Prof. K. N. Stevens	Dr. Margaret Bullowa	Dr. D. H. Klatt
Prof. M. Halle	Dr. H. Funkenstein	Dr. Paula Menyuk
Prof. W. L. Henke	Dr. A. W. F. Huggins	Dr. J. S. Perkell
Prof. A. V. Oppenheim	Dr. Emily F. Kirstein	A. R. Kessler

Graduate Students

T. Baer	R. M. Mersereau	M. R. Sambur
R. E. Crochiere	B. Mezrich	J. S. Siegel
D. E. Dudgeon	M. R. Portnoff	V. W. Zue

A. SECOND EXPERIMENT ON TEMPORALLY SEGMENTED SPEECH

1. Introduction

When a continuous speech message is switched alternately to the left and right ear of a listener, or is periodically interrupted, its intelligibility passes through a minimum when the bursts of speech reaching the listener last approximately 150 ms.[1,2] A similar effect occurs with "temporally segmented" speech,[3] which is made by splicing silent intervals into a recording of continuous speech, and hence dividing it into "speech intervals" separated from each other by "silent intervals." The advantage of using material segmented in this way is that the durations of speech and silent intervals can be independently varied, without discarding any of the speech.

In an earlier experiment[3] use was made of this advantage. Speech interval duration was varied between 30 ms and 500 ms, and silent intervals were varied independently, by making each silent interval a constant multiple of the preceding speech interval. Three experimental tapes were made, identical in all respects except that silent intervals were 40% as long, the same length, and 80% longer than the speech intervals, respectively. The intelligibility of the speech was measured as a function of speech segment duration for each tape. Each of the three functions showed a v-shaped minimum of intelligibility. Furthermore, when intelligibility was plotted against the duration of the speech intervals, the "longer duration" side of the three minima fell on a single function. When the same data were plotted against the duration of the silent intervals, the "shorter duration" side of the three minima fell on a (different) single function. This suggests that the dip in intelligibility is actually caused by two separate effects: (a) the probability that a speech interval will be recognized in isolation decreases as its duration decreases, and (b) the probability that signal parameters can be followed from

*This work was supported by the U.S. Navy Office of Naval Research (Contract N00014-67-A-0204-0064), the National Institutes of Health (Grant 5 RO1 NS04332-09), and the National Science Foundation (Grant GK-31353).

one speech interval to the next increases as the duration of the intervening silent interval decreases. The experiment reported here tested the foregoing hypothesis, using temporally segmented speech.

2. Method

Two sets of nine 100-word passages were each temporally segmented in two ways. In the first, speech interval duration was varied in 9 logarithmic steps from 31 ms to 500 ms, with silent interval duration held constant at 200 ms (that is, long enough to prevent the listener from following the speech parameters across the silent interval). In the second set of passages, speech intervals were held constant at 63 ms (that is, short enough to be unintelligible in isolation), and silent interval duration was increased from 31 ms to 500 ms. The durations of the constant silent intervals for the first tape, and of the constant speech intervals for the second tape, were chosen by extrapolation from the results of the previous experiment.[3] Eight subjects shadowed one set of passages that was temporally segmented with constant silent intervals, and the other set temporally segmented with constant speech intervals. The order of presentation and the two sets of passages were appropriately counterbalanced.

3. Results

The pooled shadowing scores are shown in Fig. X-1, and are in close agreement with the predictions of the hypothesis that is being tested. When the duration of the silent

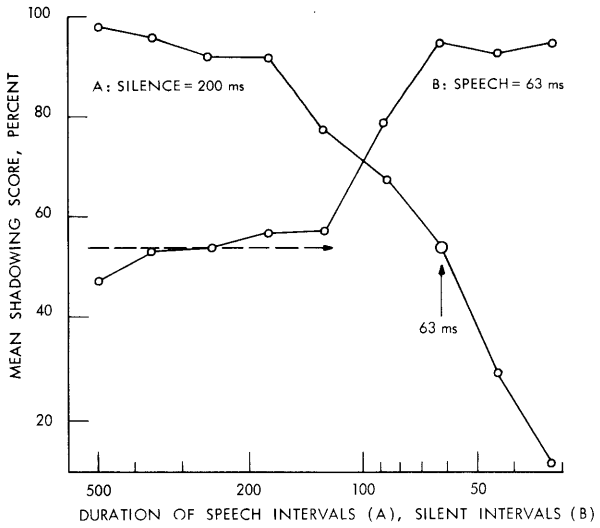

Fig. X-1. A. Mean shadowing score plotted against speech interval duration, with silent intervals held constant at 200 ms.
B. Mean shadowing score plotted against silent interval duration, with speech intervals held constant at 63 ms.

interval is held constant at 200 ms (curve A), speech intelligibility falls off from close to 100%, when speech segments last 200 ms, to ~10% when the speech segments last 31 ms. When speech segment duration is held constant at 63 ms (curve B), silent intervals shorter than 60 ms do not significantly affect intelligibility. But as silent interval duration is increased from 60 ms to 125 ms, intelligibility decreases to an asymptote at approximately 55% – which is exactly the value that would be predicted from the other function (A) for 63-ms speech intervals presented "in isolation." This provides an unexpected but striking confirmation of the hypothesis that is being tested. Note also that the 200-ms duration selected as the constant silent interval, when the effect of varying the duration of the speech intervals was being measured, is long enough to be on the asymptotic part of curve B, so that the speech segments can be considered to have been presented in isolation.

4. Discussion

Clearly, the intelligibility of temporally segmented speech depends upon the durations of both the speech and the silent intervals. The following conclusions can be drawn:

(a) When the speech intervals are separated by more than approximately 120 ms of silence, their intelligibility is not greatly affected by the duration of the silent interval. That is, the listener can extract no more information from the sequence of speech intervals than he could by combining the information obtainable from each speech segment in isolation.

(b) When silent intervals are shortened to less than 120 ms, the ability of the listener to extract information from the sequence of speech intervals rapidly improves, until the inserted silent intervals have little effect when their duration is less than ~60 ms.

(c) When silent intervals are long enough so that each speech interval is effectively presented in isolation, intelligibility decreases continuously as the speech interval duration is decreased below ~200 ms.

As a test of the adequacy of the proposed explanation of the intelligibility function for temporally segmented speech, an attempt was made to predict the results of the previously reported experiment[3] from the results of the present experiment. The model proposes that the probability that a particular speech interval will be recognized (p(S)) is the sum of the probability that it can be recognized in isolation ($p(S_{ii})$) and the probability that it can be combined with an adjacent speech interval despite the intervening silence (p(B)) if it was not identified in isolation. That is to say,

$$p(S) = p(S_{ii}) + (1-p(S_{ii})) \cdot p(B).$$

Appropriate values for the probabilities were read directly from Fig. X-1, for the speech and silent interval durations used in the earlier experiment,[3] when silent intervals were made equal to, and 80% longer than, the preceding speech intervals,

respectively. The results predicted from the foregoing simple model can be compared with the observed results in Fig. X-2. As can be seen, the agreement is very good, in part because the same speech passages, segmented in the same places, were used (wherever possible) in the two experiments.

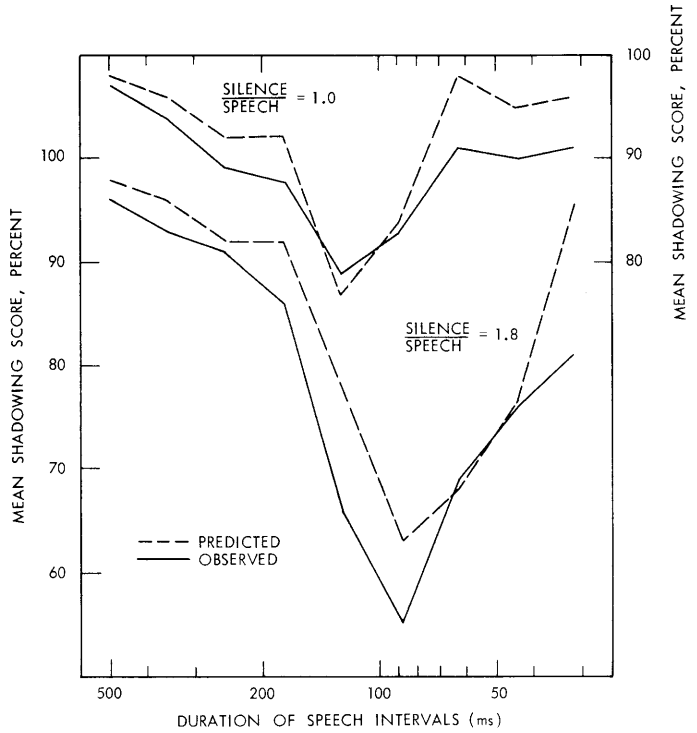

Fig. X-2. Comparison of the observed mean shadowing scores with scores predicted by the model for the intelligibility of temporally segmented speech in which the silent interval durations are a constant multiple of the speech interval durations. Shadowing scores are plotted against the duration of the speech intervals.

The close similarity between results with temporally segmented speech and with alternated speech[1,2] suggests that the explanation proposed for temporally segmented speech probably covers alternated speech as well. It might be possible to extend the model to explain interrupted speech also, but here a third variable is involved, the "discard" interval. It would be possible to design temporally segmented speech experiments to measure the effect of varying only the discard interval, while holding speech intervals and silent intervals constant. If these experiments were successful, it might then be possible to predict the intelligibility of speech produced by various schemes for time-compression of speech.

A. W. F. Huggins

References

1. E. C. Cherry and W. K. Taylor, "Some Further Experiments upon the Recognition of Speech, with One and with Two Ears," J. Acoust. Soc. Am. 26, 554-559 (1954).
2. A. W. F. Huggins, "Distortion of the Temporal Pattern of Speech: Interruption and Alternation," J. Acoust. Soc. Am. 36, 1055-1064 (1964).
3. A. W. F. Huggins, "The Perception of Temporally Segmented Speech," Quarterly Progress Report No. 103, Research Laboratory of Electronics, M.I.T., October 15, 1971, pp. 126-129; Proc. VIIth International Congress of Phonetic Sciences (Mouton, The Hague, in press).

B. SOME CONSIDERATIONS IN THE USE OF LINEAR PREDICTION FOR SPEECH ANALYSIS

1. Introduction

Recently, the analysis of speech by means of a technique referred to as linear prediction, predictive coding, or least-squares inverse filtering has received considerable attention.[1-5] This technique is directed toward modeling a sequence as the output of an all-pole digital filter. When the sequence to be modeled is specified over the domain of all integers n, there is a well-defined formulation of the technique. When only a segment of the sequence is available, however, which is always the case in practice, there are several formulations of the technique that are closely related but have important differences. One objective of this report is to summarize these differences and their implications.

When the sequence of data to be modeled is of finite length and, over the interval for which it is specified, corresponds exactly to the unit-sample response of an all-pole filter, the parameters of the model obtained by using linear prediction may be nonunique. If the data correspond closely, but not exactly, to the unit-sample response of an all-pole filter, then the solution will be unique, but the unit-sample response of the resulting filter may be considerably different from the data and small changes in the data will result in large changes in the parameters of the model and its unit-sample response. A second objective of this report is to discuss this property of the technique.

2. Formulation of the Linear-Prediction Problem

We shall consider the formulation of the technique for two problems. In problem A the data are specified for all n, and in problem B only a finite segment of the data is available.

a. Problem A

Consider a sequence s(n) defined for all n and for which $s(n) = 0$ for $n < 0$. We seek

an all-pole filter with transfer function $\hat{S}(z) = \dfrac{a_o}{1 - \sum_{k=1}^{p} a_k z^{-k}}$ such that its unit-sample response $\hat{s}(n)$ approximates $s(n)$. From the form of $\hat{S}(z)$, $\hat{s}(n)$ for $n > 0$ is given by

$$\hat{s}(n) = \sum_{k=1}^{p} a_k \hat{s}(n-k). \tag{1}$$

In the linear-prediction technique we define a predicted value of $s(n)$, denoted by $\tilde{s}(n)$, as

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{2}$$

and choose the parameters a_k to minimize the error $\mathscr{E}$ defined as

$$\begin{aligned} \mathscr{E} &= \sum_{n=1}^{\infty} |s(n)-\tilde{s}(n)|^2 \\ &= \sum_{n=1}^{\infty} \left| s(n) - \sum_{k=1}^{p} a_k s(n-k) \right|^2 . \end{aligned} \tag{3}$$

We note that the sum on n excludes the origin because $s(o)$ depends only on a_o and cannot affect the result of the minimization.

By setting $\partial\mathscr{E}/\partial a_j$ to zero for $j = 1, 2, \ldots, p$, we arrive at the following set of linear equations:

$$\sum_{k=1}^{p} a_k \sum_{n=1}^{\infty} s(n-j)\, s(n-k) = \sum_{n=1}^{\infty} s(n)\, s(n-j) \qquad j = 1, 2, \ldots, p. \tag{4}$$

In matrix notation,

$$\underline{\Phi}\, \underline{a} = \underline{\psi}, \tag{5}$$

where

$$\phi_{ij} = \sum_{n=1}^{\infty} s(n-i)\, s(n-j). \tag{6}$$

$$\psi_j = \phi_{oj}.$$

It can be shown that the matrix is symmetric and Toeplitz; that is,

$$\phi_{ij} = \phi_{ji}$$

and

$$\phi_{i+1,j+1} = \phi_{ij}.$$

Therefore the solution of Eq. 5 is computationally straightforward and efficient.[6]

We shall examine some of the properties of this solution.

i. If s(n) is indeed the unit-sample response of an all-pole filter $s(n) = \sum_{k=1}^{q} b_k s(n-k)$, then with $q \leq p$ the procedure leads to the unique solution $a_k = b_k$ for $k = 1, 2, \ldots, q$, and $a_k = 0$ for $k > q$.

ii. It can be shown that the solution $\hat{S}(z)$ corresponds to a stable filter.[7]

iii. It can be shown that the error $\mathscr{E}$ is monotonic with p.

iv. Let us define the autocorrelation function of $\hat{s}(n)$ as

$$\hat{R}_j = \sum_{n=0}^{\infty} \hat{s}(n)\, \hat{s}(n-j) \tag{7}$$

and the autocorrelation function of s(n) as

$$R_j = \sum_{n=0}^{\infty} s(n)\, s(n-j). \tag{8}$$

It can be shown that $\hat{R}_j$ and R_j are related by $\hat{R}_j = [\hat{R}_o/R_o]R_j$ for $j = 1, 2, \ldots, p$.

v. Minimizing Eq. 3 is equivalent to minimizing

$$\mathscr{E} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{S(e^{j\omega})}{\hat{S}(e^{j\omega})} - 1 \right|^2 d\omega, \tag{9}$$

where

$$\hat{S}(z) = \frac{a_o}{1 - \sum_{k=1}^{p} a_k z^{-k}}.$$

Therefore, the error criterion can be interpreted in a slightly different manner. Let u(n) denote the output of the filter $S^{-1}(z)$ when it is excited by s(n). The linear-prediction technique then corresponds to determining the $\{a_k\}$ such that u(n) is best approximated by a unit sample at n = 0.

From Eq. 9, we see that the error is dependent on the ratio of $S(e^{j\omega})$ vs $\hat{S}(e^{j\omega})$. It is also clear that the minimization depends on both the magnitude and phase of $S(e^{j\omega})$. Since $\hat{S}(e^{j\omega})$ can be shown to have minimum phase (that is, all poles and zeros are inside the unit circle), this procedure will work best when $S(e^{j\omega})$ is also minimum-phase. Heuristically, we can argue this in the following way: Since $S(z)$ is stable, we shall concern ourselves only with the zeros of $S(z)$. If $S(z)$ has zeros inside the unit circle, each of these zeros (excluding those at the origin) can be approximated by multiple poles by Taylor series expansion, and the approximation will improve as we increase p, the order of $\hat{S}(z)$. This suggests that if s(n) is minimum-phase, the error asymptotically goes to zero as p increases. This is no longer true, however, if $S(z)$ is not minimum-phase. Consequently it would be expected that if $S(z)$ is not minimum-phase, the error will not asymptotically approach zero as p increases.

b. Problem B

In this case we consider a finite segment of data of length N which we wish to model as the output of an all-pole filter. Typically, this problem has been formulated in two ways.

Formulation I

The data are multiplied by a window w(n) and the N data points are numbered from n = 0 to n = (N-1). The window is of duration N so that multiplication of the data by the window results in a sequence s'(n) which is zero for n < 0 and n > (N-1). The sequence s(n) in problem A is then taken as the sequence s'(n). In this case most of the results of problem A remain unchanged, although the sum over n is now finite. The matrix is again Toeplitz and the set of equations can be solved efficiently.

This formulation is sometimes referred to as the autocorrelation method, since the matrix $\underline{\Phi}$ is an autocorrelation matrix of s(n), as in problem A. Empirically, it has been found that for small N it is necessary to multiply s(n) by a smooth window rather than simply to truncate it, in order to minimize the end effects.[3,4]

Formulation II

No assumption is made about the data outside the interval on which they are given. Specifically, the first p values of the data are taken as initial conditions and it is assumed that with n = 0 denoting the beginning of the interval on which the data are given, the input to the all-pole filter is zero for $p \leq n < N$. The error $\mathcal{E}$ is defined as

$$\mathcal{E} = \sum_{n=p}^{N-1} \left| s(n) - \sum_{k=1}^{p} a_k s(n-k) \right|^2 ,$$

where s(n) denotes the data. To minimize the error, we set $\frac{\partial \mathcal{E}}{\partial a_j} = 0$ for $j = 1, 2, \ldots, p$ and arrive at the set of equations

$$\sum_{k=1}^{p} a_k \sum_{n=p}^{N-1} s(n-k)\, s(n-j) = \sum_{n=p}^{N-1} s(n)\, s(n-j) \qquad j = 1, 2, \ldots, p$$

or, in matrix notation, $\underline{\Phi}\,\underline{a} = \underline{\psi}$, where

$$\phi_{ij} = \sum_{n=p}^{N-1} s(n-i)\, s(n-j) \qquad i = 0, 1, \ldots, p$$

and

$$\psi_j = \phi_{oj} \qquad j = 1, 2, \ldots, p.$$

This formulation is sometimes referred to as the covariance method. The resulting matrix $\underline{\Phi}$ is still symmetric but no longer Toeplitz. In fact,

$$\phi_{i+1,j+1} = \sum_{n=p}^{N-1} s(n-i-1)\, s(n-j-1) = \sum_{n=p-1}^{N-2} s(n-i)\, s(n-j)$$

$$= \phi_{ij} - s(N-1-i)\, s(N-1-j) + s(p-1-i)\, s(p-1-j). \qquad (10)$$

The last terms in Eq. 10 can be considered an end-effect correction.

Both formulations have been used by researchers, hence it is appropriate to compare their efficiency. This is shown in Table X-1. Formulation I has the following advantages. Theoretically, the stability of the resulting filter $\hat{S}(z)$ is guaranteed, (although this is not true for implementation with finite word-length computation). Increasing p from p_o to $p_o + 1$ involves only one additional iteration; therefore, it is easy to set an error threshold to select the appropriate value of p. On the other hand, Formulation II has the advantages that scaling is relatively simple for fixed-point implementation, and the computation can be carried out in-place. It has also been pointed out that the square-root method of solving the resulting set of linear equations is numerically very stable.[9]

Table X-1. Computational efficiency of Formulations I and II.

	Formulation I	Formulation II
Matrix	Toeplitz (Can Be Solved by Levinson's Method[6])	Symmetric (Can Be Solved by Square-Root Method[8])
Storage (data)	N	N
matrix equation	$4p+4$	$(p^2+3p)/2$
window	$2N$	0
Computation		
multiplies (windowing)	N	0
(compute ϕ_{ij})	$pN-p^2$	$pN+p$
(solve matrix equation)	$2p^2+5p-6$	$(p^3+9p^2+2p)/6$
divides (or inverse)	$2p$	p
square-roots	0	p

3. Application to Sequences Closely Approximating the Response of an All-Pole Filter

The linear-prediction method is most suitable for sequences that can be closely approximated as the response of an all-pole filter. Typically, the linear-prediction technique is used to determine the parameters of the all-pole filter, and spectral analysis or resynthesis is carried out by using these parameters to generate an approximation, $\hat{s}(n)$, to the data.

For problem A we assumed, by virtue of Eq. 1 and the fact that the data are zero for $n < 0$, that the input was a unit sample at $n = 0$. Thus, to generate $\hat{s}(n)$ from the parameters, we excite the all-pole filter with a unit sample. For problem B, the autocorrelation method outlined in Formulation I suggests the same procedure, since the product of the data and the window is treated as in problem A. For Formulation II in problem B, we made the assumption that for $p \leq n < N$ the filter input is zero. It is useful to consider the result of applying linear prediction to data that do correspond exactly to the output of an all-pole filter so that the data s(n) satisfy the relationship

$$s(n) = \sum_{k=1}^{q} b_k s(n-k) \tag{11}$$

on a specified interval, and we choose to estimate s(n) by

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{12}$$

on that interval. For problem A the interval is $1 \leq n \leq \infty$ and if $q \leq p$, $\hat{s}(n) = s(n)$, and the coefficients a_k will be equal to b_k for $1 \leq k \leq q$ and zero for $k > q$. For problem B, the autocorrelation formulation (Formulation I) will not in general give $\hat{s}(n) = s(n)$, and the a_k will not equal the b_k because the infinite duration sequence corresponding to the data multiplied by the window no longer satisfies the relationship of Eq. 11. With $q \leq p$ the covariance method will always give $\hat{s}(n) = s(n)$ over the interval. When $p = q$ the a_k are uniquely determined and specify a system whose unit-sample response is $\hat{s}(n) = s(n)$. For $p > q$, if the data satisfy Eq. 11 for $p \leq n < N$ but not for $0 \leq n < p$, then we conjecture that the a_k are uniquely determined. Moreover, the unit-sample response of the all-pole filter, $\hat{s}(n)$, will equal $s(n)$ only if $s(n)$ corresponds to the unit-sample response of an all-pole filter. The fact that the specified data $s(n)$ satisfy the relationship of Eq. 11 does not require it to be the unit-sample response of an all-pole filter but only that it be the response to an input which is zero for $p \leq n < N$. Now let us consider the case for which the data satisfy Eq. 11 for $0 \leq n < N$; that is, all of the specified data including the initial conditions satisfy Eq. 11. With $q \leq p$ the covariance method will always give $\hat{s}(n) = s(n)$ over the interval. When $p = q$ the a_k are uniquely determined and specify a system whose unit-sample response is $\hat{s}(n) = s(n)$. When $p > q$, however, the $(p \times p)$ matrix $\underline{\Phi}$ is of rank q, which gives a $p-q$ parameter family of solutions for the a_k. For each solution vector $\underline{a}$ in the family of solutions, $\hat{s}(n) = s(n)$, but one and only one of these solutions specifies a system whose unit-sample response is $\hat{s}(n) = s(n)$. This solution is, of course, the one for which all of the a_k vanish when $k > q$.

Consider the following simple example. Let $s(n)$ be exactly the unit-sample response of the filter

$$S(z) = \frac{1}{1 - a z^{-1}}.$$

Thus $s(n) = a^n u_{-1}(n)$. Suppose $s(n)$ is estimated by the second-order linear predictor

$$\tilde{s}(n) = a_1 s(n-1) + a_2 s(n-2).$$

We choose to minimize the mean-square error on the interval $[n_o, n_o+1]$, using $s(n_o-1)$ and $s(n_o-2)$ as starting values (Formulation II). If $n_o > 1$, then the equation $\underline{\Phi}\,\underline{a} = \underline{\psi}$ is

$$\begin{bmatrix} a^2 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a^3 \\ a^2 \end{bmatrix}.$$

Clearly, the matrix $\underline{\Phi}$ is singular and the general solution for $\underline{a}$ can be expressed as any particular solution $\underline{a}^o$ added to any linear combination of vectors spanning the null space of $\underline{\Phi}$. We choose $\underline{a}^o = \begin{bmatrix} a \\ 0 \end{bmatrix}$ which is the particular solution for which $\hat{s}(n)$ is the unit-sample response of the filter

$$H_o(z) = \frac{1}{1 - a z^{-1}}.$$

The general solution is given by

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + c \begin{bmatrix} -a \\ a^2 \end{bmatrix},$$

where c is an arbitrary constant. The solution $\underline{a}$ thus lies along the line $a_2 = -a a_1 + a^2$ in the $a_1 a_2$ plane. To illustrate a different solution in the solution space, suppose $c = 1$. Then $a_1 = 0$, $a_2 = a^2$, and $\hat{s}(n) = s(n)$ is generated by the filter

$$H_1(z) = \frac{1}{1 - a^2 z^{-2}} = \frac{1}{(1 - a z^{-1})(1 + a z^{-1})},$$

excited not by a unit sample, but by the sequence

$$x(n) = u_o(n) + a u_o(n-1).$$

We see that the pole of $H_1(z)$ at $z = -a$ is canceled by the zero at $z = -a$ of the input sequence.

That the predictor coefficients are not, in general, unique is not surprising. The linear-prediction problem as formulated seeks to determine a difference equation, whose solution approximates a given sequence on some interval. If this difference equation is associated with a linear system, we see that there is nothing "built into" the formulation of the problem that specifies the initial conditions of the system, that is, exactly how the system was excited. All that is required by the present formulation of the problem is that the input to the system vanish over the interval on which s(n) is being predicted. Hence the multiplicity of solutions may be interpreted as resulting from the fact that different systems with different inputs can produce identical outputs.

In practice, we are not generally interested in applying linear prediction to a sequence that is exactly the output of an all-pole filter of unknown order. Thus we do not expect the covariance matrix to be singular (when it is singular it can be dealt with by choosing $p = \text{rank } \underline{\Phi}$). We are interested, however, in applying linear prediction to sequences which may be modeled approximately as the output of an all-pole filter. If the sequence

s(n) closely approximates the output of an all-pole filter (as speech often does) the covariance matrix $\underline{\Phi}$ will be ill-conditioned, that is, almost singular. Although a unique solution does exist, it appears to be very sensitive to small perturbations in the data. In particular, if a small amount of noise is added to the data which, according to the previous discussion, results in a family of solutions, the resulting solution may be close to any one of the solutions in the family, and as small perturbations are introduced into the data, the resulting solution may change radically. A consequence is that if the order of the predictor is too large, and the data are close to the unit-sample response of an all-pole filter, as is often assumed to be the case in speech analysis, the unit-sample response of the all-pole filter specified by the linear-prediction parameters and its Fourier transform, may not approximate the data very closely, although the output of the filter resulting from another unspecified input will.

4. Summary and Conclusion

We have attempted to point out the major differences between the various formulations of the linear-prediction problem and discuss a set of important issues related to linear prediction. We have seen that there are generally two different methods of formulating this problem; one requires s(n) to be zero outside the domain of minimization, and the other does not. The autocorrelation method provides a good match to the spectrum, but this is by no means an indication of its superiority over the covariance method.

The uniqueness of the linear-prediction solution is a very important issue. Our experience indicates that with additive noise injected, the system does not always converge to a desirable answer. What perceptual effect this has on speech synthesis is still unclear. We hope to answer this question better after experimental speech synthesis.

M. R. Portnoff, V. W. Zue, A. V. Oppenheim

References

1. B. S. Atal and Suzanne L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am. 50, 637-665 (1971).

2. C. J. Weinstein and A. V. Oppenheim, "Predictive Coding in a Homomorphic Vocoder," IEEE Trans., Vol. AU-19, pp. 243-248, 1971.

3. J. Markel, "Formant Trajectory Estimation from a Linear Least-Squares Inverse Filter Formulation," Monograph No. 7, Speech Communications Research Laboratory, Inc., Santa Barbara, California, 1971.

4. J. I. Makhoul, "Aspects of Linear Prediction in the Spectral Analysis of Speech," paper presented at IEEE-AFCRL 1972 International Conference on Speech Communication and Processing, Boston, Massachusetts, April 24-26, 1972.

5. V. W. Zue, "Speech Analysis by Linear Prediction," Quarterly Progress Report No. 105, Research Laboratory of Electronics, M.I.T., April 15, 1970, pp. 133-142.

6. N. Levinson, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," Appendix B in N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series (John Wiley and Sons, Inc., New York, 1949).

7. U. Grenander and G. Szegö, Toeplitz Forms and Their Applications (University of California Press, Berkeley, 1958), p. 40.

8. V. N. Faddeeva, Computational Methods of Linear Algebra, English Translation by C. D. Benster (Dover Publications, Inc., New York, 1959), pp. 81-85.

9. J. H. Wilkinson, Rounding Errors in Algebraic Processes (Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1963), pp. 117-118.