Proceedings of The Institute of Acoustics

THE POLE-ZERO ANALYSIS OF SPEECH BY GENERALISED LINEAR PREDICTION THEORY.

H.C. LONGUET-HIGGINS,
LABORATORY OF EXPERIMENTAL PSYCHOLOGY,
UNIVERSITY OF SUSSEX.

The linear predictive coding of speech (Atal & Hanauer 1972, Markel & Gray 1976) is based on the assumption that a sufficiently short sequence of speech samples s_n may be described as the convolution of an unpredictable source e_n with an effectively constant impulse response sequence t_n. Equation (1) expresses this assumption in the language of z^{-n} transforms:

$$S(z) = T(z)E(z). \qquad (1)$$

It is commonly further assumed that the transfer function T(z) has poles but no zeros: but a more general assumption is that

$$T(z) = B(z)/A(z), \qquad (2)$$

where

$$A(z) = 1 - \sum_{1}^{p} a_k z^{-k} = \prod_{1}^{p}(1-z_k/z) \qquad (3)$$

and

$$B(z) = 1 - \sum_{1}^{q} b_k z^{-k} = \prod_{1}^{q}(1-z'_k/z). \qquad (4)$$

It will now be shown that the poles $(z_1,\ldots,z_p)$ and the zeros $(z'_1,\ldots,z'_q)$ may all be found from the first p+q samples of the autocorrelation function, if the values of p and q can be reliably estimated in advance.

It is convenient to represent the autocorrelation function of the speech by the power series

$$R(z) = S(z)S(1/z) = \sum_{-\infty}^{\infty} r_j z^{-j} \qquad (5)$$

where, in terms of the speech samples s_n,

$$r_j = r_{-j} = \sum_{0}^{\infty} s_n s_{n+j} \ : \qquad (6)$$

the assumption of an uncorrelated source e_n then enables one to write

$$R(z) = \lambda T(z)T(1/z), \qquad (7)$$

where λ is a constant. Multiplying both sides of this equation by A(z) one obtains

THE POLE-ZERO ANALYSIS OF SPEECH BY GENERALISED LINEAR PREDICTION THEORY.

$$A(z)R(z) = B(z)T(1/z), \tag{8}$$

in which, on the right-hand side, there are no powers of z less than z^{-q}. Equating coefficients of z^{-n}, and introducing $a_o = -1$, we thus obtain

$$\sum_{o}^{p} a_k r_{n-k} = 0 \ (n > q), \tag{9}$$

which may be written in the more explicit notation

$$\begin{bmatrix} r_q, & r_{q-1}, & \dots, & r_{q-p+1} \\ r_{q+1}, & r_q, & & \vdots \\ \vdots & & & \\ r_{q+p-1}, & & \dots, & r_q \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_{q+1} \\ \vdots \\ r_{q+p} \end{bmatrix} . \tag{10}$$

It should be remarked that if p and q are <u>overestimates</u> of the numbers of poles and zeros, then the matrix on the left-hand side will be singular.

Having determined the coefficients a_k and thence the poles z_k, one may proceed to locate the zeros z_k' by the following method, assuming that they all lie within the unit circle.

The identity

$$C(z) = A(z)R(z)A(1/z) = \lambda B(z)B(1/z) \tag{11}$$

shows that the zeros of B(z) are included among the roots of

$$C(z) = c_o + c_1(z+z^{-1}) + \dots + c_q(z^q+z^{-q})=0, \tag{12}$$

where

$$c_n = \sum_{o}^{p}\sum_{o}^{p} a_j r_{n-j+k} a_k. \tag{13}$$

Hence, by finding the roots of (12) and choosing from each pair of reciprocal roots that one which lies within the unit circle we obtain the zeros of B(z), which are the required zeros of T(z). The coefficients b_k are then immediately found from equation (4).

Equation (14),

$$A(z)S(z) = B(z)E(z), \tag{14}$$

written in the form

$$s_n - \sum_{1}^{p} a_k s_{n-k} = e_n - \sum_{1}^{q} b_k e_{n-k}, \qquad (15)$$

enables one to determine the source samples from the speech samples, or vice versa. This equation represents a generalisation of the standard equation of linear prediction theory.

The above analysis has been successfully applied to the recovery of the zeros as well as the poles of synthetic speech segments, generated either from a sequence of clicks or from a white noise source, and weighted with a hanning window. For real speech the results are less satisfactory, partly because of the difficulty of choosing a suitable value of q. Further studies are in progress, directed towards improving the estimation of q and enhancing the stability of the computations.

References

1. B.S. ATAL and S. HANAUER 1972 J. Acoust. Soc. Amer. 50, 637-655. Speech Analysis and Synthesis by Linear Prediction of the Speech Wave.

2. S.D. MARKEL and A.H. GRAY Jr. 1976. Linear Prediction of Speech. Springer-Verlag, Berlin/Heidelberg/New York.