Linear Prediction

# 1 Transfer Function Models

Discrete signal , considered as the output of a system with input , such that

|  |  |
| --- | --- |
|  | (1) |

where and . The parameters of the hypothesised system are , and gain . Taking the z-transform of equation (1) gives the transfer function of the system,

|  |  |
| --- | --- |
|  | (2) |

This is a general pole-zero model. For an all-pole model and for an all-zero model, . The all-zero model is also known as a moving average (MA) model and the all-pole model is known as an autoregressive (AR) model. The general pole-zero model is also known as an autoregressive moving average (ARMA) model.

# 2 Least Squares Estimate

Consider a system described by an all-pole model,

|  |  |
| --- | --- |
|  | (3) |

The transfer function is

|  |  |
| --- | --- |
|  | (4) |

The goal is to estimate the model parameters , known as the predictor coefficients, and the gain given a particular input signal . Assume that output is unknown (which is common). The signal can be estimated from past samples,

|  |  |
| --- | --- |
|  | (5) |

The error, or residual, is

|  |  |
| --- | --- |
|  | (6) |

The total squared error is given by

|  |  |
| --- | --- |
|  | (7) |

We want to obtain the system parameters that result in the least total squared error. To do this, set

|  |  |
| --- | --- |
|  | (8) |

From equations (7) and (8) we obtain

|  |  |
| --- | --- |
|  | (9) |

for . Equations (9) are known as the ‘normal equations’. For any definition of , we get a set of equations with unknowns that can be solved to obtain the predictor coefficients that minimise . The minimum total squared error can be shown to be

|  |  |
| --- | --- |
|  | (10) |

# 3 Autocorrelation Method

Assume the time duration is infinite, that is . Equations (9) and (10) reduce to

|  |  |
| --- | --- |
|  | (11) |
|  | (12) |

where the autocorrelation is defined as

|  |  |
| --- | --- |
|  | (13) |

Note that is an even function, that is . The coefficients form the autocorrelation matrix, which is a symmetric Toeplitz matrix[[1]](#footnote-1). Equation (11) in matrix form is

|  |  |
| --- | --- |
|  | (14) |

Equation (1.14) can be solved using various methods, including

* Gauss reduction/elimination
* Crout reduction method
* Cholesky decomposition method
* **Durbin’s recursive procedure**

# 4 Covariance Method

Unlike in the autocorrelation method, assume the signal is defined for a finite time interval, that is . Equations (9) and (10) reduce to

|  |  |
| --- | --- |
|  | (15) |
|  | (16) |

where the covariance is

|  |  |
| --- | --- |
|  | (17) |

Similar to the autocorrelation method, this method uses to form a covariance matrix, which is symmetric. Unlike the autocorrelation matrix, the covariance matrix does not have equal diagonal terms (i.e. not a Toeplitz matrix).

# 5 Frequency Domain Formulations

Recall that the error is given by

|  |  |
| --- | --- |
|  | (18) |

Taking the z-transform,

|  |  |
| --- | --- |
|  | (19) |

where and are the z-transform of and respectively, and is the inverse filter. Recall that the transfer function is

|  |  |
| --- | --- |
|  | (20) |

Applying Parseval’s Theorem, the total error to be minimised is

|  |  |
| --- | --- |
|  | (21) |

where .

The power spectrum of is given by , where . This allows the energy to be minimised to be expressed as

|  |  |
| --- | --- |
|  | (22) |

Minimising the error by applying the procedure applied in equation (8). The result is identical to the autocorrelation normal equations (9), but with the autocorrelation obtained from the signal spectrum by an inverse Fourier transform,

|  |  |
| --- | --- |
|  | (23) |

# 6 Linear Predictive Spectral Matching

Let be an estimate of the power spectrum based on the all-pole model,

|  |  |
| --- | --- |
|  | (24) |

From equations (19), the power spectrum is

|  |  |
| --- | --- |
|  | (25) |

Comparing (24) and (25), the error power spectrum is modelled by a flat spectrum equal to . This means that the actual error signal e, is being approximated by another signal that has a flat spectrum, such as a unit impulse, white noise, or any other signal with a flat spectrum. The filter is sometimes known as a ‘whitening filter’ since it attempts to produce an output signal , that is white (i.e., has a flat spectrum).

The total error can be expressed as

|  |  |
| --- | --- |
|  | (26) |

This allows the problem of linear prediction to be restated as follows. Given some spectrum, we wish to model it by another spectrum such that the integrated ratio between the two spectra as in (1.26) is minimized.

Consider two spectra and   with autocorrelation coefficients and respectively. The gain factor is obtained by equating the total energy in the two spectra, i.e. . Increasing the order of the model increases the range over which , resulting in a better fit of to . When , for all and the two spectra become identical . This implies that any spectrum can be approximated arbitrarily closely using an all-pole model.

Equation (24) can be rewritten as

|  |  |
| --- | --- |
|  | (27) |

where

|  |  |
| --- | --- |
|  | (28) |

is the autocorrelation of the inverse filter . As , the transfer function is given by

|  |  |
| --- | --- |
|  | (29) |

where is the minimum phase sequence corresponding to .

# 7 Long-term Predictors

For quasi-periodic signals such as speech, the error signal can be modelled as

|  |  |
| --- | --- |
|  | (30) |

where is the fundamental pitch period, are the long-term prediction coefficients and is the prediction error of the long-term filter which is a completely random signal with a white spectrum,

|  |  |
| --- | --- |
|  | (31) |

Minimising the total squared error results in

|  |  |
| --- | --- |
|  | (32) |

An alternative to modelling the short-term and long term predictions separately is to combine the short and long term predictors into a single model,

|  |  |
| --- | --- |
|  | (33) |

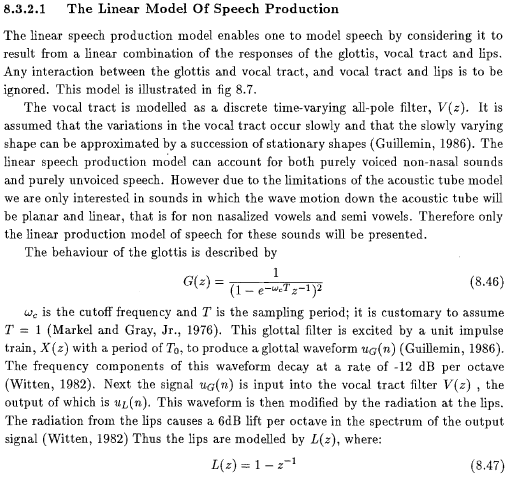
For more info: ‘Advanced Digital Signal Processing and Noise Reduction’ Ch.8 (S. Vaseghi 1996).

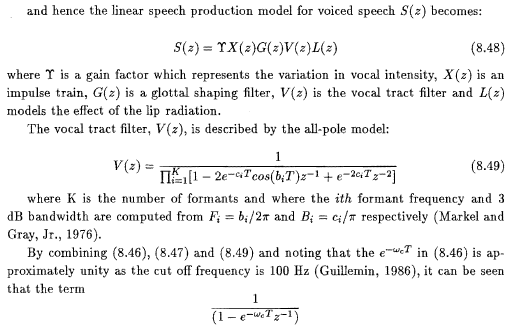
# Relevant Resources

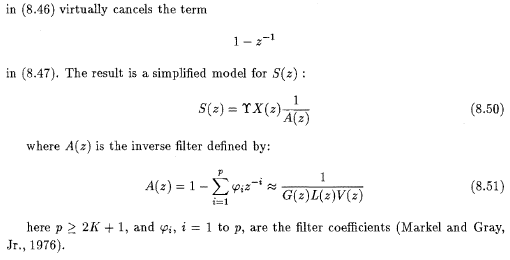
* ‘Linear Prediction: A Tutorial Review’ (J Makhoul 1975)
* ‘Advanced Digital Signal Processing and Noise Reduction’ Ch.8 (S. Vaseghi 1996)
* ‘Linear Prediction of Speech’ (J.D. Markel and A.H. Gray 1976)

A Detailed Approach

From Catherine’s Ph.D thesis,







Note: For consistency with above section, replace with and replace with .

# Justification of the Linear Prediction Method

Recall the signal estimate is

|  |  |
| --- | --- |
|  | (34) |

Taking the z-transform,

|  |  |
| --- | --- |
|  | (35) |

Where is the linear prediction filter (Markel and Gray Jr. 1976),

|  |  |
| --- | --- |
|  | (36) |

Recall the error signal is

|  |  |
| --- | --- |
|  | (37) |

Taking the z-transform,

|  |  |
| --- | --- |
|  | (38) |

Comparing (8.50) from thesis and (38), we can see that if we set , then and hence the linear prediction model of speech results in an equivalent linear production model of speech (Markel and Gray Jr. 1976).

# Long Term Analysis from paper

For all voiced sounds (e.g. vowels), the short-term error/excitation signal , is characterised by a periodicity of samples, describing the voice fundamental pitch. can therefore be expressed as

|  |  |
| --- | --- |
|  |  |

where is the fundamental pitch period, ranging in a set of all possible human pitches. is a gain term and is a wide-band noise component. An estimation of is

|  |  |
| --- | --- |
|  |  |

therefore

|  |  |
| --- | --- |
|  |  |

The parameter is obtained by minimizing the cost function

|  |  |
| --- | --- |
|  |  |

where is approximated as . When is obtained, the long-term prediction error can be obtained via equation (10).

# 3. Proposed Features

The features proposed in this paper are the prediction error energy and prediction gain for both short-term and long-term analyses, which are defined as

|  |  |
| --- | --- |
|  |  |

These quantities are computed for each window and define the vectors

|  |  |
| --- | --- |
|  |  |

where is the total number of windows. A boxcar window of length 0.025ms is used. Feature vector is comprised of mean, standard deviation, minimum and maximum values of /,

|  |  |
| --- | --- |
|  |  |

The final feature vector is obtained by repeating the linear prediction analysis with different prediction orders ,

|  |  |
| --- | --- |
|  |  |

In the proposed implementation, and .

Classification of speech signals into synthetic or bona fide is done using a ML approach, classifying the proposed features using random forest or support vector machine (SVM) methods.

The Linear Model for Speech

The linear model for speech production involves considering the behavior of the glottis, vocal tract and lips as independent filters in the z-domain. As such, any interactions between the glottis, lips and vocal tract are disregarded. The filter corresponding to the behavior of the glottis is

|  |  |
| --- | --- |
|  | (39) |

where is the cutoff frequency and is the sampling period. Typically, it is assumed that (Markel and Gray Jr. 1976). The vocal tract can be described by a filter given by

|  |  |
| --- | --- |
|  | (40) |

where is the number of formants and and , where is the -th formant frequency and is the corresponding 3 dB bandwidth (Markel and Gray Jr. 1976). Finally, the behavior of the lips is captured in the z-domain using a filter,

|  |  |
| --- | --- |
|  | (41) |

As the effect of the glottis, vocal tract and lips are considered to be independent processes, the input signal is thought to pass through each filter successively. The input signal is … Once this signal has passed through the glottal filter , the resulting output is the glottal waveform (Guilleman 1986). The frequency components of decays at a rate of -12 dB/octave (Witten 1982). This signal is passed through the vocal tract filter to produce an output , which is modified by the radiation at the lips, which causes a 6 dB/octave lift (Witten 1982) to produce the output speech signal .

If and are the respective z-transforms of the input signal and the output signal , then the speech production model can be stated as

|  |  |
| --- | --- |
|  | (42) |

where is a gain factor which captures the variation in the intensity of the voice and is the voice production filter, given by

|  |  |
| --- | --- |
|  | (43) |

As is typically small, the approximation can be used to cancel the zero in the numerator, leading to the all-pole approximation of ,

|  |  |
| --- | --- |
|  | (44) |

This means that can be modelled by an all pole filter of the form

|  |  |
| --- | --- |
|  | (45) |

where is the inverse filter, given by

|  |  |
| --- | --- |
|  | (46) |

where is the prediction order (Markel and Gray Jr. 1976).

# Linear Prediction of Speech

Taking the inverse z-transform of equation (42) shows that speech can be modelled as an autoregressive process,

|  |  |
| --- | --- |
|  | (47) |

As a general rule, the prediction order should be at least to ensure accurate prediction.

The error is

|  |  |
| --- | --- |
|  | (1) |

This can be written in matrix form as

|  |  |
| --- | --- |
|  | (1) |

where  is a vector containing all of the speech samples,

|  |  |
| --- | --- |
|  | (1) |

is a matrix with row containing the past samples used to predict the -th element of ,

|  |  |
| --- | --- |
|  | (1) |

and is a vector of the predictor coefficients,

|  |  |
| --- | --- |
|  | (1) |

The predictors resulting in the least square error is found by minimizing the L2 norm of the error vector,

|  |  |
| --- | --- |
|  | (1) |

resulting in the normal equations

|  |  |
| --- | --- |
|  | (1) |

Now, define as the autocorrelation matrix. It can be shown that and that the element in row and column is

|  |  |
| --- | --- |
|  | (1) |

Moreover, is a Toeplitz matrix, meaning that elements along lines parallel to the diagonal are equal. The normal equations in expanded form are

|  |  |
| --- | --- |
|  | (1) |

1. A Toeplitz matrix is one where all the elements along each diagonal are equal. [↑](#footnote-ref-1)