Dublin Institute of Technology
ARROW@DIT

Conference papers

Audio Research Group

2008-01-01

Linear Prediction: The Problem, its Solution and Application to Speech

Alan O'Cinneide
Dublin Institute of Technology, alan.ocinneide@dit.ie

David Dorran
Dublin Institute of Technology, david.dorran@dit.ie

Mikel Gainza
Dublin Institute of Technology, mikel.gainza@dit.ie

Follow this and additional works at: http://arrow.dit.ie/argcon

Part of the Signal Processing Commons

Recommended Citation

O'Cinneide, A., Dorran, D., Gainza, M.: Linear Prediction: The Problem, its Solution and Application to Speech. DIT Internal Technical Report.

This Working Paper is brought to you for free and open access by the Audio Research Group at ARROW@DIT. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

Linear Prediction

The Technique, Its Solution and Application to Speech

Alan Ó Cinnéide

August 2008

Supervisors: Drs. David Dorran, Mikel Gainza and Eugene Coyle

Contents

1 Overview and Introduction **2**

2 Linear Systems: Models and Prediction **2**

2.1 Linear System Theory 3

2.2 All-Pole Linear Prediction Model 5

2.3 Solutions to the Normal Equations 7

2.3.1 The Autocorrelation Method 7

2.3.2 The Covariance Method 9

2.3.3 Comparison of the Two Methods 9

3 Linear Prediction of Speech **10**

3.1 Human Speech Production: Anatomy and Function 10

3.2 Speech Production as a Linear System 12

3.2.1 Model Limitations 13

3.3 Practical Considerations 14

3.3.1 Prediction Order 14

3.3.2 Closed Glottal Analysis 15

4 Examples **16**

4.1 Human Speech: Voiced Vowel 16

4.2 Human Speech: Unvoiced Fricative 16

4.3 Trumpet 17

1 Overview and Introduction

Linear prediction is a signal processing technique that is used extensively in the analysis of speech signals and, as it is so heavily referred to in speech processing literature, a certain level of familiarity with the topic is typically required by all speech processing engineers. This paper aims to provide a well-rounded introduction to linear prediction, and so doing, facilitate the understanding of the technique. Linear prediction and its mathematical derivation will be described, with a specific focus on applying the technique to speech signals. It is noted, however, that although progress in linear prediction has been driven primarily by speech research, it involves concepts that prove useful to digital signal processing in general.

First to be discussed within the paper are general linear time-invariant systems, along with its theory and mathematics, before moving into a general description of linear prediction models. The equations that yield one variant of linear prediction coefficients are derived and the methods involved to solve these equations are then briefly discussed. Different interpretations of the equations yield slightly different results, and these differences will be explained.

A section focussing specifically on the linear prediction of speech then begins. The anatomical process of speech production is described, followed by an introduction to a theoretical linear model of the process. The limitations of applying the linear prediction model to speech are described, and comments are also given concerning certain practicalities that are specific to the linear prediction of speech.

The paper concludes with an implementation of linear prediction using two different types of signal. It is hoped that the balance between theory and practical will allow the reader for easy assimilation of this technique.

2 Linear Systems: Models and Prediction

Linear prediction [11, 7, 9] is a technique of time series analysis, that emerges from the examination of linear systems. Using linear prediction, the parameters of a such a system can be determined by analysing the systems inputs and outputs. Makhoul [7] says that the method first appeared in a 1927 paper

on sun-spot analysis, but has since been applied to problems in neuro-physics, seismology as well as speech communication.

This section will review linear systems and, elaborating upon them, derives the mathematics of linear prediction.

2.1 Linear System Theory

A linear system is such that produces its output as a linear combination of its current and previous inputs and its previous outputs [13]. It can be described as time-invariant if the system parameters do not change with time. Mathematically, linear time-invariant (LTI) systems can be represented by the following equation:

$$y(n) = \sum_{j=0}^{q} b_j x(n-j) - \sum_{k=1}^{p} a_k y(n-k) \tag{1}$$

This is the general difference equation for any linear system, with output signal y and input signal x, and scalars b_j and a_k, for $j = 1 \ldots q$ and $k = 1 \ldots p$ where the maximum of p and q is the order of the system. The system is represented graphically in figure 1.

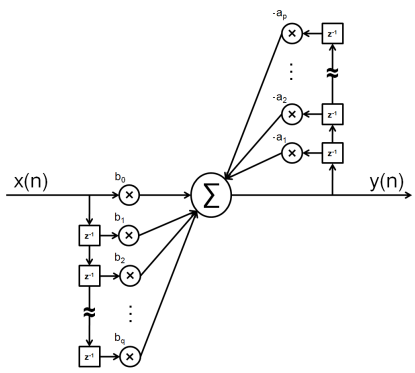

Figure 1: A graphical representation of the general difference equation for an LTI system.

By re-arranging equation (1) and transforming into the Z-domain, we can

reveal the transfer function $H(z)$ of such a system:

$$y(n) + \sum_{k=1}^{p} a_k y(n-k) = \sum_{j=0}^{q} b_j x(n-j)$$

$$\sum_{k=0}^{p} a_k y(n-k) = \sum_{j=0}^{q} b_j x(n-j) \quad \text{where } a_0 = 1$$

$$\sum_{k=0}^{p} a_k z^{-k} Y(z) = \sum_{j=0}^{q} b_j z^{-j} X(z)$$

$$\Rightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{j=0}^{q} b_j z^{-j}}{\sum_{k=0}^{p} a_k z^{-k}} \tag{2}$$

The coefficients of the input and output signal samples in equation (1) reveal the poles and zeros of the transfer function.

Linear prediction follows naturally from the general mathematics of linear systems. As the system output is defined as a linear combination of past samples, the system's future output can be predicted if the scaling coefficients b_j and a_k are known. These scalars are thus also known as the predictor coefficients of the system [9].

The general linear system transfer function gives rise to three different types of linear model, dependent on the form of the transfer function $H(z)$ given in equation (2) [9, 7].

- When the numerator of the transfer function is constant, an *all-pole* or *autoregressive* (AR) model is defined.
- The *all-zero* or *moving average* model assumes that the denominator of the transfer function is a constant.
- The third and most general case is the *mixed pole/zero* model, also called the *autoregressive moving-average* (ARMA) model, where nothing is assumed about the transfer function.

The all-pole model for linear prediction is the most widely studied and implemented of the three approaches, for a number of reasons. Firstly, the input signal, which is required for ARMA and all-zero modelling, is oftentimes

an unknown sequence. As such, they are unavailable for use in our derivations. Secondly, the equations derived from the all-pole model approach are relatively straight-forward to solve, contrasting sharply with the nonlinear equations dervied from ARMA or all-zero modelling. Finally, and perhaps the most important reason why all-pole modelling is the preferred choice of engineers, many real world applications, including most types of speech production, can be faithfully modeled using the approach.

2.2 All-Pole Linear Prediction Model

Following from the linear system equation (1), one can formulate the equations necessary to determine the parameters of an all-pole linear system, the so-called linear prediction normal equations. First, following on from the all-pole model (see Figure 2), a linear prediction estimate $\hat{y}$ at sample number n for the output signal y by a pth order prediction filter can be given by:

$$\hat{y}(n) = -\sum_{k=1}^{p} a_k y(n-k) \tag{3}$$

The error or residue between the output signal and its estimate at sample n

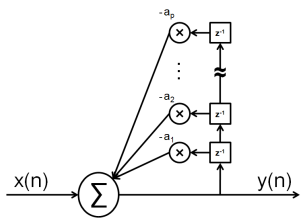

Figure 2: A graphical representation of an all pole linear system, where the output is a linear function of scaled previous outputs and the input.

can then be expressed as the difference between the two signals.

$$e(n) = y(n) - \hat{y}(n) \tag{4}$$

The total squared error for an as-of-yet unspecified range of signal samples is given by the following equation:

$$\begin{aligned} E &= \sum_n [e(n)]^2 \\ &= \sum_n [y(n) - \hat{y}(n)]^2 \\ &= \sum_n [y(n)]^2 - 2 \cdot y(n) \cdot \hat{y}(n) + [\hat{y}(n)]^2 \end{aligned} \tag{5}$$

Equation (5) gives a value indicative of the energy in the error signal. Obviously, it is desirous to choose the predictor coefficients so that the value of E is minimised over the unspecified interval. The optimal minimising values can be determined through differential calculus, i.e. by obtaining the derivative of equation 5 with respect to each predictor coefficient and setting that value equal to zero.

$$\begin{aligned} &\frac{\partial E}{\partial a_k} = 0 \quad \text{for } 1 \leq k \leq p \\ \Rightarrow &\frac{\partial}{\partial a_k}(\sum_n([y(n)]^2 - 2 \cdot y(n) \cdot \hat{y}(n) + [\hat{y}(n)]^2)) = 0 \\ &-2\sum_n y(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) + 2\sum_n \hat{y}(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) = 0 \\ &\sum_n y(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) = \sum_n \hat{y}(n) \cdot \frac{\partial}{\partial a_k}\hat{y}(n) \\ &\frac{\partial}{\partial a_k}\hat{y}(n) = -y(n-k) \quad \ldots \text{from equation (3)} \\ \Rightarrow &\sum_n y(n) \cdot -y(n-k) = \sum_n \hat{y}(n) \cdot -y(n-k) \\ &-\sum_n y(n) \cdot y(n-k) = \sum_n (-\sum_{i=1}^{p} a_i y(n-i)) \cdot -y(n-k) \\ &-\sum_n y(n) \cdot y(n-k) = \sum_{i=1}^{p} a_i \sum_n y(n-i) \cdot y(n-k) \end{aligned} \tag{6}$$

For the sake of brevity and future utility, a *correlation function* ϕ is defined. The expansion of this summation describes what will be called the *correlation matrix*.

$$\phi(i,k) = \sum_n y(n-i) \cdot y(n-k) \tag{7}$$

Substituting the correlation function into equation (6) allows it to be written

more compactly:

$$-\phi(0,k) = \sum_{i=1}^{p} a_i \phi(i,k) \tag{8}$$

The derived set of equations are called the normal equations of linear prediction.

2.3 Solutions to the Normal Equations

The limits on the summation of the total squared energy were omitted from equation (5) so as to give their selection special attention. The section will show that two different but logical summation intervals lead to a two different sets of normal equations and result in different predictor coefficients.

Given sufficient data points and appropriate limits, the normal equations define p equations with p unknowns which can be solved by any general simultaneous linear equation solving algorithms, e.g. Gaussian elimination, Crout decomposition, etc. However, certain limits lead to matrix redundancies and allow for efficient solutions that can significantly reduce the computational load.

2.3.1 The Autocorrelation Method

The autocorrelation method of linear prediction minimises the error signal over all time, from $-\infty$ to $+\infty$. When dealing with finite digital signals, the signal is windowed such that all samples outside the interval of interest are taken to be 0 (see Figure 3). If the signal is non-zero from 0 to $N-1$, then the resulting error

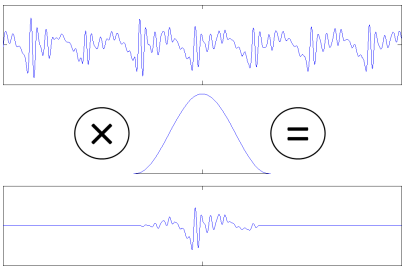

Figure 3: Windowing a signal by multiplication with an appropriate function, in this case a Hanning window.

signal will be non-zero from 0 to $N-1+p$. Thus, summing the total energy

over this interval is mathematically equivalent to summing over all time.

$$\begin{aligned} E &= \sum_{n=-\infty}^{\infty} [e(n)]^2 \\ &= \sum_{n=0}^{N-1+p} [e(n)]^2 \end{aligned} \tag{9}$$

When these limits are applied to equation (7), a useful property emerges. Because the error signal is zero outside the analysis interval, the correlation function of the normal equations can be identically expressed in a more convenient form.

$$\begin{aligned} \phi_{auto}(i,k) &= \sum_{n=0}^{N-1+p} y(n-i) \cdot y(n-k) \quad 1 \leq i \leq p \quad 1 \leq k \leq p \\ &= \sum_{n=0}^{N-1+(i-k)} y(n) \cdot y(n+(i-k)) \quad 1 \leq i \leq p \quad 1 \leq k \leq p \end{aligned}$$

This form of the correlation function is simply the short-time autocorrelation function of the signal, evaluated with a lag of $(i-k)$ samples. This fact gives this method of solving the normal equations its name.

The implications of this convenience is such that the correlation matrix defined by the normal equations exhibits a double-symmetry that can exploited by a computer algorithm. Given that $a_{i,j}$ is the member of the correlation matrix on the ith row and j^{th} column, the correlation matrix demonstrates:

- standard symmetry, where $a_{i,j} = a_{j,i}$,

$$\begin{pmatrix} a_{1,1} & a_{2,1} & a_{3,1} & \cdots & a_{m,1} \\ a_{2,1} & a_{2,2} & a_{3,2} & \cdots & a_{m,2} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{m,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}$$

- Toeplitz symmetry, where $a_{i,j} = a_{i-1,j-1}$.

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,m} \\ a_{2,1} & a_{1,1} & a_{1,2} & \cdots & a_{m,2} \\ a_{3,1} & a_{2,1} & a_{1,1} & \cdots & a_{m,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{1,1} \end{pmatrix}$$

These redundancies mean that the normal equations can be solved using the Levinson-Durbin method, an recursive procedure that greatly reduces the computational load.

2.3.2 The Covariance Method

In contrast with the autocorrelation method, the covariance method of linear prediction minimises the total squared energy only over the interval of interest.

$$E = \sum_{n=0}^{N-1} [e(n)]^2 \tag{10}$$

Using these limits, an examination of the equation (7) reveals that the signal values required for the calculation extend beyond the analysis interval (see Figure 4).

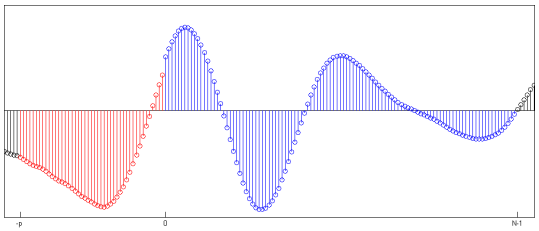

Figure 4: The covariance method require $-p$ samples (shown here in red) beyond the analysis interval from 0 to $N-1$ (shown in blue).

$$\phi_{covar}(i,k) = \sum_{n=0}^{N-1} y(n-i) \cdot y(n-k) \quad 1 \leq i \leq p \quad 1 \leq k \leq p \tag{11}$$

Samples are required from $-p$ to $N-1$. The resulting correlation matrix exhibits standard symmetry, but unlike the matrix defined by the autocorrelation mehthod, the matrix does not demonstrate Toeplitz symmetry. This means that a different method must be used to solve the normal equations, such as Cholesky decomposition or the square-root method.

2.3.3 Comparison of the Two Methods

Each of these solutions to the linear prediction normal equations has its own strengths and weaknesses; determining which is more advantageous to use is greatly determined by the signal being analysed. When analysis signals are long,

the two different solutions are virtually identical. Because of the greater redundancies in the matrix defined by the autocorrelation method, it is slightly easier to compute [9]. Experimental evidence indicates that the covariance method is more accurate for periodic speech sounds [1], while the autocorrelation method performs better for fricative sounds [9].

3 Linear Prediction of Speech

In order for linear prediction to apply to speech signals, the speech production process must closely adhere to the theoretical framework established in the previous sections. This section reviews the actual physical process of speech production and discusses the linear model utilised by speech engineers.

3.1 Human Speech Production: Anatomy and Function

Following Figure 5, the vast majority of human speech sounds are produced in the following manner [3]. The lungs initiate the speech process by acting as the bellows that expels air up into the other regions of the system. The air pressure is maintained by the intercostal and abdominal muscles, allowing for the smooth function of the speech mechanisms. The air that leaves the lungs then enters into the remaining regions of the speech production system via the trachea. This organ system, consisting of the lungs, trachea and interconnecting channels, is known as the pulmonary tract. The turbulent air stream is driven up the trachea into the larynx. The larynx is a box-like apparatus that consists of muscles and cartilage. Two membranes, known as the vocal folds, span the structure, supported at the front by the thyroid cartilage and at the back by the arytenoid cartilages. The arytenoids are attached to muscles which enable them to approximate and separate the vocal folds. Indeed, the principal function of the larynx, unrelated to the speech process, is to seal the trachea by maintaining the vocal folds closed. This has the dual benefit of being able to protect the pulmonary tract and permit the build up of pressure within the chest cavity necessary for certain exertions and coughing [9].

The space between the vocal folds is called the glottis. A speech sound is classified as voiced or voiceless depending on the glottal behaviour as air passes

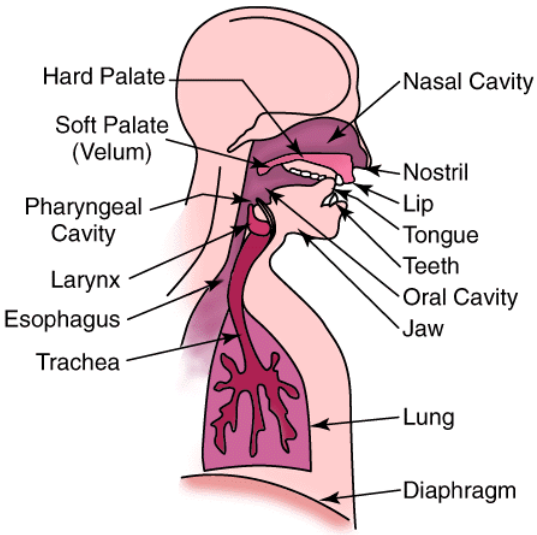

Figure 5: The human speech production system. Image taken from http://cobweb.ecn.purdue.edu/~ee649/notes/physiology.html.

through it. According to the myeloelastic-aerodynamic theory of phonation [2], the vibration of the vocal folds results from the interaction between two opposing forces. Approximated folds are forced apart by rising subglottal air pressure. As air rushes through the glottis, the suction phenomenon known as the Bernoulli effect is observed. This effect due to decreased pressure across the constriction aperture adducts the folds back together. The interplay between these forces results in vocal fold vibration, producing a voiced sound. This phonation has a fundamental frequency directly related to the frequency of vibration of the folds. During a voiceless speech sound, the glottis is kept open and the stream of air continues through the larynx without hindrance. The resulting glottal excitation waveform exhibits a flat frequency spectrum.

The phonation from the larynx then enters the various chambers of vocal tract: the pharynx, the nasal cavity and the oral cavity. The pharynx is the chamber stemming the length of the throat from the larynx to the oral cavity. Access to the nasal cavity is dependent on the position of the velum, a piece of soft tissue that makes up the back of the roof of the mouth. For the production of certain phonemes, the velum descends, coupling the nasal cavity with the

other chambers of the vocal tract.

The fully realised phone is radiated out of the body via either the lips or the nose or both. This is a continuous process, during which the state and configuration of the system's constituents alter and change dynamically with the thoughts of the speaker.

3.2 Speech Production as a Linear System

The acoustic theory of speech production assumes the speech production process to be a linear system, consisting of a source and filter [11]. This model captures the fundamentals of the speech production process described in the previous section: a source phonation modulated in the frequency domain by a dynamically changing vocal tract filter, Figure 6. According to the source-filter theory,

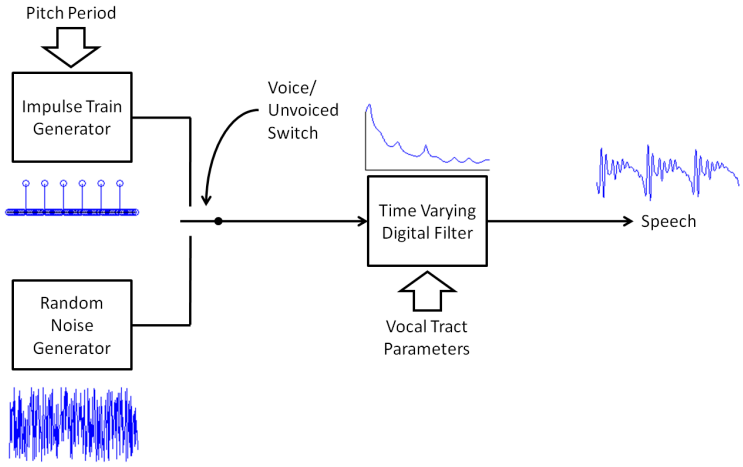

Figure 6: The simplified speech model proposed by the acoustic theory of speech production.

short-time frames of speech can be characterised by identifying the parameters of the source and filter.

Glottal source. The source signal is one of two states: a pulse train of a certain fundamental frequency for voiced sounds and white noise for unvoiced sounds. This two-state source fits reasonably well with true glottal behaviour, though moments of mixed excitation cannot be represented well.

Vocal tract filter. The vocal tract is parameterised by its resonances, which are called *formants*[1]. All acoustic tubes have natural resonances, the parameters of which are a function of its shape.

Though the vocal tract changes its shape, and thus its resonances, continuously with running speech, it is not unreasonable to assume it static over short-time intervals of the order of 20 milliseconds. Thus, speech production can be viewed as a LTI system and linear prediction can be applied to it.

3.2.1 Model Limitations

In truth, the speech production system is known to have some nonlinearity effects and the glottal source and vocal tract filter are not completely de-coupled. In other words, the acoustic effects of the vocal tract has been noticed to modulate depending on the behaviour of the source in ways that linear systems cannot fully describe. Additionally the vocal tract deviates from the behaviour of an all-pole filter during the production of certain vocal sounds.

System linearity. Linear systems by definition assume that inputs to the system have no effect on the system's parameters [13]. In the case of the speech production process, this means that the vibratory behaviour of the glottis has no bearing on the formant frequencies and bandwidths - an assumption which is sometimes violated [2]. Especially in the situation where the pitch of the voice is high and the centre frequency first formant low, an excitation pulse can influence the decay of the previous pulse.

All pole model. The described method of linear prediction works on the assumption that the frequency response of the vocal tract consists of poles only. This supposition is acceptable for most voiced speech sounds, but is not appropriate for nasal and fricative sounds. During the production of these types of utterances, zeros are produced in the spectrum due to the trapping of certain frequencies within the tract. The use of a model lacking representation of zeros in addition to poles should not cause too much concern, as if p is of high enough order, the all pole model is sufficient for almost all speech sounds [11].

[1] The word formant comes from the Latin verb *formāre* meaning "to shape".

Despite these limitations, all-pole linear prediction remains a highly useful technique for speech analysis.

3.3 Practical Considerations

Analysing speech signals using linear prediction requires a couple of provisos to achieve the optimal results. These considerations are discussed in this section.

3.3.1 Prediction Order

The choice of prediction order is an important one as it determines the characteristics of the vocal tract filter. Should the prediction be too low, key areas of resonance will be missed as there are insufficient poles to model them - if the prediction order is too high, source specific characteristics, e.g. harmonics, are determined (see Figure 7). Formants require two complex conjugate poles to

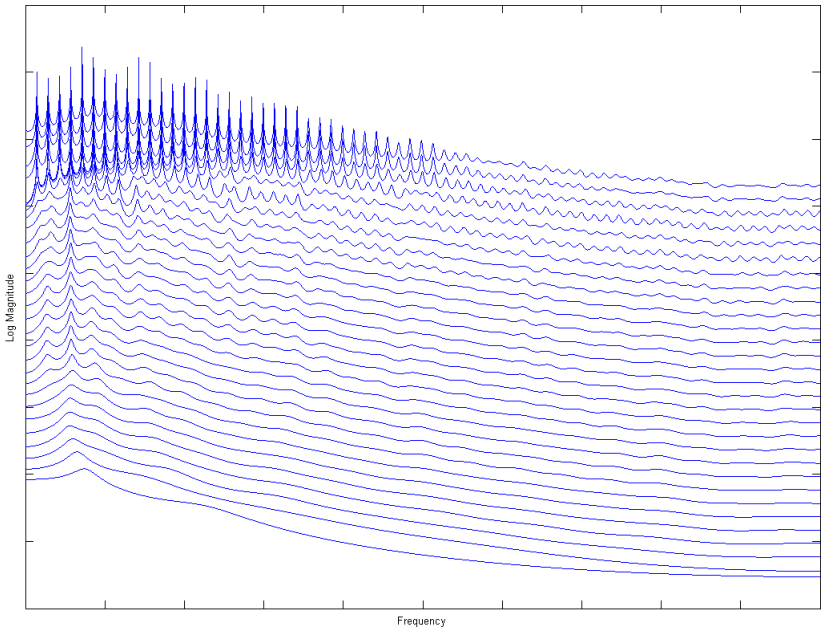

Figure 7: The spectral envelopes of a trumpet sound, as determined by linear prediction analyses, each successively increased prediction orders.

characterise correctly. Thus, the prediction order should be twice the number of formants present in the signal bandwidth. For a vocal tract of 17 centimetres

long, there is an average of one formant per kilohertz of bandwidth.

Where p represents the prediction order and f_s the signal's sampling frequency, the following formula is used as a general rule of thumb [9]:

$$p = \frac{f_s}{1000} + \gamma \tag{12}$$

The value of γ, described in the literature as a "fudge factor", necessary to compensate for glottal roll-off and predictor flexibility, is normally given as 2 or 3.

3.3.2 Closed Glottal Analysis

The general consensus of the speech processing community is that linear predictive analysis of voiced speech should be confined to the closed glottal condition [5]. Indeed, it has been shown that closed phase covariance method linear prediction yields better formant tracking and inverse filtering results that than other pitch synchronous and asynchronous methods [1]. During the glottal closed phase the the signal represents a freely decaying oscillation, theoretically ensuring the absence of source-filter interaction and thus better adhering to the assumptions of linear prediction [15].

Some voice types are unsuited to this type of analysis [10]. In order to obtain a unique solution to the normal equations, a critical minimum of signal samples must exist related to the signal's bandwidth. High-pitched voices are known to have closed phases that are too short for analysis purposes. Other voices, particularly breathy voices, are known to exhibit continuous glottal leakage.

There are numerous methods used to determine the closed glottal interval. The first attempts to do so typically used special laboratory techniques, such as electroglottography [4], recorded simultaneously with the digital audio. More recently, efforts have focused on ascertaining the closed glottal interval through the analysis of the recorded speech signal [14, 6]. Some of these techniques has met significant success, particularly the DYPSA algorithm of Naylor et al. [8] that successfully identifies the closed glottal instant in more than 90% of cases (Figure 8).

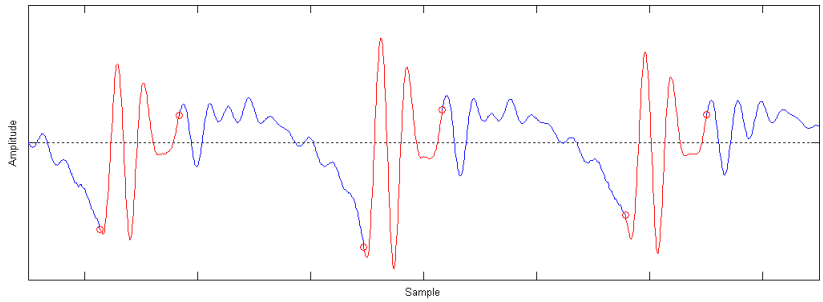

Figure 8: A speech waveform with the closed glottal interval highlighted in red and delimited by circles. These regions represent the instants of glottal closing and opening respectively, as identified by the DYPSA algorithm.

4 Examples

Within this section, some implementations of linear prediction are given, along with all the practical considerations taken for the analysis.

4.1 Human Speech: Voiced Vowel

A voice sample of a male voice, recorded at a sampling rate of 44.1 kHz, was analysed. The signal is the voiced vowel sound $/a/$. As it is periodic, covariance method linear prediction during the closed glottal phase yield the most accurate formant values.

The signal was first processed by the DYPSA algorithm to determine the closed glottal interval, which underwent covariance analysis. The order of the prediction filter, calculated according to formula (12), was determined to be 46, see figure 9.

4.2 Human Speech: Unvoiced Fricative

An unvoiced vocal sample was also analysed. This segment, sampled at a rate of 9 kHz, is taken from the TIMIT speech database and is of the fricative sound $/\int/$, the *sh* found in both "*sh*ack" and "ca*sh*", as pronounced by an American female.

As autocorrelation linear prediction analysis performs better with unvoiced sounds, that method was implemented with a filter of prediction order of 11,

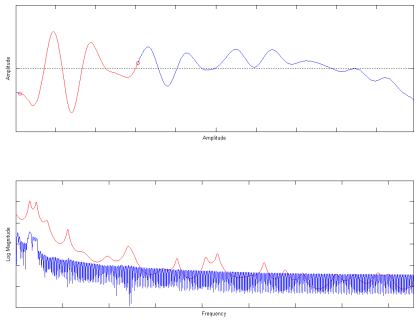

Figure 9: **Top:** time-domain representation of $/a/$ sound. **Below:** The sound's spectrum and spectral envelope as determined by covariance method linear prediction analysis.

see figure 10.

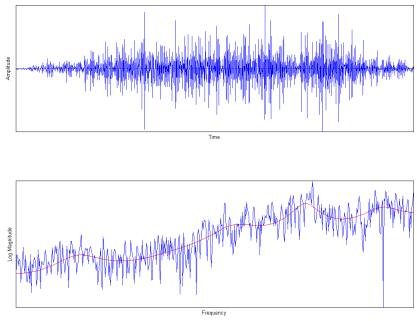

Figure 10: **Top:** time-domain representation of $/\int/$ sound. **Below:** The sound's spectrum and spectral envelope as determined by autocorrelation method linear prediction analysis.

4.3 Trumpet

Though this report has primarily concerned itself with the linear prediction of speech, linear prediction also has applications for musical signal processing [12]. Certain instrumental sounds, such as brass instruments, exhibit strong formant structure that lend themselves well to modelling through linear prediction.

In this example, given in figure 11, a B♭ trumpet sample was analysed, playing the E♭5. Trumpets are known to exhibit 3 formants, indicating a prediction

order of 6 is required to determine all the resonances. As the signal is periodic, covariance method linear prediction analysis is performed.

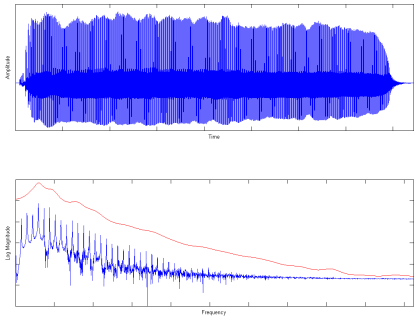

Figure 11: **Top:** time-domain representation of a trumpet playing the note E♭5. **Below:** The sound's spectrum and spectral envelope as determined by covariance method linear prediction analysis.

References

[1] S. Chandra and Lin Wen. Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis. *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, 22(6):403–415, 1974.

[2] D. G. Childers. *Speech Processing and Synthesis Toolboxes*. John Wiley & Sons, Inc., New York, 2000.

[3] Michael Dobrovolsky and Francis Katamba. Phonetics: The sounds of language. In William O'Grady, Michael Dobrovolsky, and Francis Katamba, editors, *Contemporary Linguistics: An Introduction*. Addison Wesley Longman Limited, Essex, 1997.

[4] A. Krishnamurthy and D. Childers. Two-channel speech analysis. *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, 34(4):730–743, 1986.

[5] J. Larar, Y. Alsaka, and D. Childers. Variability in closed phase analysis of speech. volume 10, pages 1089–1092, 1985.

[6] Changxue Ma, Y. Kamp, and L. F. Willems. A frobenius norm approach to glottal closure detection from the speech signal. *Speech and Audio Processing, IEEE Transactions on*, 2(2):258–265, 1994.

[7] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[8] Patrick A. Naylor, Anastasis Kounoudes, Jon Gudnason, and Mike Brookes. Estimation of glottal closure instants in voiced speech using the dypsa algorithm. *Audio, Speech and Language Processing, IEEE Transactions on [see also Speech and Audio Processing, IEEE Transactions on]*, 15(1):34–43, 2007.

[9] T. W. Parsons. *Voice and speech processing.* McGraw-Hill New York, 1987.

[10] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds. Modeling of the glottal flow derivative waveform with application to speaker identification. *Speech and Audio Processing, IEEE Transactions on*, 7(5):569–586, 1999.

[11] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals.* Prentice-Hall Signal Processing Series. Prentice-Hall Inc., Engelwood Cliffs, New Jersey, 1978.

[12] Curtis Roads. *The Computer Music Tutorial.* MIT Press, Boston, Massachusetts, 1996.

[13] A. V. Oppenheim R. W. Schafer. *Digital Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey, 1975.

[14] R. Smits and B. Yegnanarayana. Determination of instants of significant excitation in speech using group delay function. *Speech and Audio Processing, IEEE Transactions on*, 3(5):325–333, 1995.

[15] D. Wong, J. Markel, and Jr. Gray, A. Least squares glottal inverse filtering from the acoustic speech waveform. *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on*, 27(4):350–355, 1979.