LARYNX SYNCHRONOUS FORMANT ANALYSIS

L C Whitaker and D J B Pearce

ABSTRACT

The paper describes and evaluates the performance of several methods of larynx synchronous formant analysis for a female speaker. The techniques are based on the covariance method of LPC analysis and root solving of the coefficients to obtain the pole positions. The results obtained however, depend on both the length and position of the analysis interval. In this paper it is demonstrated that analysis over the closed phase of the larynx gives better performance than fixed frame length or analysis over the whole larynx cycle. The positions of the glottal closure have been determined using a laryngograph signal.

The improved performance is demonstrated by a better ability to follow the transient features of the signal with fewer missed or extra formants, and better formant continuity. The ability to follow formant transitions during glides (e.g. w,r,l) and in voiced segments following plosives is particularly apparent. Closed phase analysis also shows improved performance over other analyses with white noise added to the signal.

INTRODUCTION

Commonly all short-time speech analysis techniques use frame sizes between 10 ms and 40 ms. During a typical voiced frame the speech signal will be the response of the vocal tract excited at the glottis. Each frame will contain samples produced when the larynx is closed and opened. When the larynx is opened the vocal tract is coupled with the sub-glottal system which causes source and vocal tract interaction. During the closed larynx phase there is no such coupling and the vocal tract acts as a resonant cavity. The consequences of performing the analysis over framesizes greater than the closed larynx interval is that the source/tract interaction is changing during the analysis interval which may affect spectral estimates. This is avoided by doing the analysis over the closed phase since there is no excitation during this interval.

COMPARISON AND EVALUATION OF CLOSED PHASE ANALYSIS WITH OTHER METHODS

The speech data was down-sampled from 20 kHz to 10 kHz and a 10 pole LPC analysis and root-solving was performed. Three analysis intervals were used for the purpose of comparison: (1) the closed phase interval (CPA), (2) the whole larynx cycle (LCA), (3) a fixed 15 ms framemlength (FLA).

The larynx closure and openings instants were determined by means of a laryngograph (ref 2). The laryngograph measures the r.f. impedance between two electrodes placed on the neck at the larynx. This impedance is related to the lateral area of contact of the vocal folds. Larynx
closure is rapid and is defined by the maximum gradient on the rising edge of each larynx cycle. Larynx opening is estimated by thresholding the \( l_x \) at the point of closure.

EXISTENCE OF CLOSED PHASE INTERVAL

In order to demonstrate the existence of a closed phase region in speech a sample by sample analysis was used. In this method the analysis is performed over an interval usually less than the closed phase, the window is moved forward by one sample and the analysis is repeated. Figure 1 shows a sample by sample analysis of the vowel /e/ spoken by a female speaker. The diagram demonstrates that a closed phase region exists which is constant for approximately 1 ms. The effect of transition from the closed phase to the open larynx phase is an increase in the scattering of the pole positions caused by the increased influence of the excitation and sub glottal coupling. This is also observed in the acoustic signal by an increased damping of the signal envelope.

EFFECT OF FUNDAMENTAL FREQUENCY

Atal (ref 3) showed that periodic excitation is likely to introduce errors in the LPC analysis. Results showed that estimated formant frequencies were pulled towards the nearest harmonic frequency. These errors were estimated to be worse for the first formant frequency and for those speakers with high fundamental frequency. Closed phase analysis is not effected by such problems because the analysis is during an interval where there is no excitation. The effect of fundamental frequency on the 3 analyses methods is illustrated in Figure 2 for the vowel /e/ where \( f_0 \) is falling rapidly from 281-134 hz. Gross differences are observed in the first pole position between the CPA which is consistently higher in value than the LCA and FLA particularly when \( f_0 \) is high.

The fundamental frequency also effects the size of the closed phase interval due to an approximate inverse relationship. This is particularly problematic for female speakers, when at high \( f_0 \) the closed phase interval may be too small to maintain analysis accuracy. This is likely if the number of samples in the closed phase interval is smaller than twice the number of poles in the analysis. In this case the minimum closed phase interval is set at twice the number of poles.

FAST FORMANT TRANSITIONS

Comparison of CPA, LCA and FLA have been performed for 3 types of formant transitions:

1. The smooth, fairly fast formant transition typcial of lateral sounds. This is illustrated for the sentence "why are you early you owl" in Figure 3. The CPA gives improved tracking for the initial /w/ in "why". Further differences in shape and smoothness are observed between the CPA and other two methods during the lateral sound /j/ in you. The CPA also gives better resolution when the formants are close together during the word "owl" than the LCA and FLA.

2. Formant transitions due to very sharp voice onsets typical of plosive sounds. This is demonstrated for the analysis of "speech" in Figure 4 where the formant transitions at the release of the unvoiced stop
/p/ for the CPA is much more rapid than for the FLA. The differences are because the FLA is performed over several larynx cycles.

3 The transition from a nasal to a following vowel. This is shown for the word "nine" in Figure 5. For the CPA the first formant transition is abrupt whereas for the FLA and LCA there are poles present between the steady state values. Large improvements in the formant continuity and separation of the following diphthong (nasalised) are also observed for the CPA when compared to the FLA and LCA.

The results show that CPA can provide better tracking of the transitional features in speech than other techniques which used longer framesizes.

EFFECT OF NOISE ON ANALYSIS

Essential to any speech analysis technique is a capability to perform well in noisy environments. CPA should perform better than other techniques in noisy conditions because the closed phase region is typically of higher amplitude than the open phase and therefore the signal to noise ratio is improved. Figure 6 shows the analysis for the word "nine" with white noise added. The analyses shows that the noise causes some degradation for both analyses. The CPA in noise however still gives better formant continuity particularly for second formant. Closed phase analysis therefore can still perform better than FLA in noisy environments.

CONCLUSION

This paper evaluates the performance of CPA and compares its performance with LCA and FLA. The results of the comparison demonstrate clearly that CPA can provide more accurate formant estimates and better modelling of formant transitions. Further closed phase analysis also shows improved performance than FLA when white noise is added to the signal. Closed phase analysis has been shown to be a powerful method for analysing voiced speech. This work has several useful applications in particular improved modelling of the vocal tract should provide better features for phonetic recognition.

REFERENCES

2 A J Fourcin and E Abberton, "First applications of a new laryngograph", Medical and Biological Illustration, 21, 172-182
3 B S Atal, "Linear prediction of speech - Recent advances with applications to speech analysis", Bell Labs Journal
Fig 1: Sample by sample analysis for vowel segment /e/

Fig 2: Plots of F0 and F1 for CPA, LCA and FLA for vowel segment /e/

Fig 3: Analysis of the sentence "why are you early you owl?"

Fig 4: Analysis of the word "speech"

Fig 5: Analysis of word "nine"

Fig 6: Analysis of word "nine" with white noise added