Proceedings of The Institute of Acoustics

AN IMPROVED FORMANT TRACKING ALGORITHM FEATURING ADAPTIVE POLE ENHANCEMENT

C. Duncan and M. A. Jack

Centre for Speech Technology Research,
University of Edinburgh.

INTRODUCTION

The measurement of areas of spectral resonance (formants) in the spectrum of speech is essential in the identification of voiced speech sounds in recognition systems employing knowledge-based feature extraction and interpretation. Measurement of formant centre frequency is crucial in the identification of vowels, and additionally, pre- and post-vocalic consonant identification is enhanced if formant movement, particularly that of the second formant (F2), can be tracked with respect to time.

Linear prediction coding (LPC) techniques [1,2] offer a convenient method of spectral analysis of the speech waveform to allow formant estimation, but characteristically offer only moderate signal-to-noise immunity, with the limit of usefulness being at approximately 15dB signal-to-noise power ratio (SNR).

The novel technique presented in this paper employs spectral estimation using LPC techniques, but offers a substantial improvement in the signal-to-noise performance of formant frequency estimation. The objective of the technique is essentially to enhance the detection of singularities (poles) in the estimated transfer function of the vocal tract inherent in the LPC-based model. High noise immunity in formant estimation is specifically conferred by the progressive over-estimation of model order in any single analysis frame, coupled with the use of off-axis spectral estimation. An averaging filter function is applied as the final step in each analysis frame to yield noise-robust estimates of formant centre frequency.

The study of spectrum-based formant tracking in itself demands consideration of two sets of properties; those of the spectral estimator on the one hand, and those of the spectral line tracking algorithm on the other. The latter relies heavily on a knowledge of the dynamics of the speech mechanism, but the design and comparison of such algorithms is not considered in this paper. However, the examination of the performance of the spectral estimation component of the formant tracking system provides a reasonable measure of the likely performance of the overall system, particularly in its ability to maintain the integrity of any formant track and to maintain separation of tracks which may tend to merge.

In order to assess the noise robustness of the pole enhancement technique in comparison to classical LPC spectral estimators, a synthesised signal with well-controlled parameters is used. The signal-to-noise performance criteria employed in this study are based in part on difference limen related to the human perception of fine variations in the spectral structure of voiced speech [3,4].

Assessment of the technique within a formant tracking system is provided by a comparison with an advanced LPC technique [5] which also employs off-axis spectral estimation but uses a fixed model order, with both techniques sharing a common line-tracking paradigm. It is shown that the pole enhancement technique has a generally superior performance in terms of ability to separate formants which merge together, and additionally can provide a 15dB improvement in signal-to-noise immunity when used in such a tracking system.

THE POLE ENHANCEMENT TECHNIQUE

The design of the pole enhancement process is based on the well-established result in the study of two-pole resonating systems that the less the damping factor, ζ, the less the half-power bandwidth of the peak in the frequency response of the system. In addition, the centre frequency of the resonance peak, f_m, moves upwards in frequency towards the undamped natural frequency of oscillation, f_n as the damping decreases:

$$f_m = f_n\sqrt{1 - 2\zeta^2} \tag{1}$$

By analogy, the view can be taken that, as the search path in the transfer function plane is diverted from the usual Fourier transform axis so as to approach the singularity locations, the peaks associated with these poles in the resulting off-axis spectrum behave as if their damping factor decreases, i.e. their Q-factor (= centre frequency/bandwidth) and centre frequency both increase. However, in more complex resonanting systems characterised by many pole pairs, it is unlikely that the variation of any peak centre frequency will be monotonic as the transform search path approaches the associated singularity. This is due to the effect of other poles on the frequency response of the system. In general, the path of any given peak centre frequency can be expected to follow equ.(1) only at very high Q-factors, where the pole-pair associated with the peak effectively dominates the enhanced frequency response and can be considered to be decoupled from the rest of the system.

In LPC-based speech analysis, the vocal tract filter is modeled as a set of concatenated acoustic tube sections from the glottis to the lips, the amount of sections being directly dependent on the chosen LPC model order. The input excitation is assumed to be either from the periodic vibration of the vocal cords (impulse train) or from a constriction in the vocal tract causing turbulent air flow (white noise).

In general, the choice of model order is seen as having a critical effect on the performance of the spectral estimator, and usually the order is chosen to be between $p = 12$ and $p = 16$ inclusive: with a low model order, LPC analysis fails to separate formants which merge in certain speech sounds; too large a value of model order clutters the smooth spectrum with spurious peaks and is considered to degrade the signal-to-noise performance of the estimator.

The pole enhancement technique avoids the problem of masking of weak vocal tract formants by more intense formant features by employing off-axis spectral estimation. For any given analysis frame, several spectra are calculated using various values of z-transform radius. Formant candidates are then extracted from each off-axis spectrum using peak detection and parabolic interpolation. However, the improvements afforded by the technique are achieved by the deliberate progressive over-estimation of the model order with each decrease in z-transform radius. In this way, the acoustic tube model becomes progressively more likely to contain terms relating to weak formants; that is, the technique is successful at providing good spectral separation of formants which otherwise may merge. The progressive increase in model order can also be considered to aid in conferring improved noise immunity to the technique when used in conjunction with a simple averaging filter. This is since, if the speech is contaminated with noise having nonstationary spectral characteristics, then formant-like features related to additive noise can be expected to occupy different positions in the spectrum corresponding to each choice of model order. When coupled with the simultaneous use of off-axis spectral estimation, the progressive increase in model order can in contrast be expected to produce a deterministic pattern of formant movement on the short-time stationary characteristics of the spectrum, namely the true formants associated with the vocal tract. The final step in conferring high noise immunity on the formant estimation process involves the use of an averaging filter using Q-factor as a weighting criterion. The objective of this final stage is to enhance those spectral features which have remained consistent both across all pole-enhanced spectra and as model order has changed, and to attenuate formant-like features which have exhibited nonstationary spectral behaviour as z-transform radius and model order have changed.

EVALUATION OF FORMANT ESTIMATION AND TRACKING PERFORMANCE

The characterisation of the pole enhancement technique as a formant estimation algorithm is divided into two performance criteria: (1) Noise tolerance characterisation by comparing standard deviation of formant frequency estimation from nominal steady state formant frequencies against that of a standard 16^{th}-order LPC analysis in the presence of additive white noise, and (2) the ability to maintain separation of merging formants in real speech and hence maintain the integrity of formant tracking particularly with respect to the second formant.

The investigation of noise tolerance performance employed an artificially-generated speech-like signal of 100ms duration. This signal is characterised by 3 formant-like features in its short-time spectrum at nominally 500Hz, 1kHz and 2kHz with relative power amplitude of 0dB, -6dB and -12dB respectively, and each formant having a bandwidth of 100Hz. The signal is generated by exciting a simulated parallel formant synthesiser with an impulse train of frequency 150Hz ± 3%. The equivalent sample frequency for the signal is 16kHz, 12-bit resolution, with a quantisation SNR of 60dB.

In both the pole enhancement technique and standard LPC analysis, a 25.6ms Hamming-windowed analysis frame was used and the total analysis was performed by moving the window exactly one sample point at a time to give a total of 1192 analysis frames.

In the pole enhancement technique, 11 spectral estimates were produced using z-transform radii in the range $0.9 \leq r < 1$ decreasing in steps of $\delta r = 0.01$. The initial model order was $p = 12$ for $r = 1$ increasing by +2 for each step decrease in radius. The value of radius step has been chosen to some extent arbitrarily, and results have indicated that decreasing the radius much below 0.9 does not appear to improve the performance of the technique. Experimentation has found that the technique is relatively insensitive to choice of initial model order (values of between 8 and 16 were used with little observable change to the results presented here). The LPC technique used in the comparative noise tolerance characterisation employed a fixed 16th-order model with z-transform radius $r = 1$.

The evaluation process employed a statistical averaging routine to calculate average and standard deviation (with respect to the average) for each nominal formant centre frequency. This routine extracted formant values both from each signal frame and from each analysis technique on a "nearest neighbour" criterion, with the nominal formant frequencies of the artificial signal, as detailed above, being used to select the formant candidates.

The 100ms signal was contaminated by progressively more intense additive white noise in 5dB steps down to 15dB SNR and then in 1dB steps down to -6dB. The results for the standard deviation of formant frequency estimation against SNR for each technique and for formants F1 and F2 are shown in Figs. 1(a) and 1(b) for SNR values between 15dB and -4dB. The results for all curves can be considered to remain constant at SNRs above 15dB.

Initial assessment of the performance of each technique is based here on formant difference limen of 5% of nominal formant centre frequency. This gives a difference limen criterion of 25Hz for F1 and 50Hz for F2. For formant F1, the results indicate that the limit of usefulness for the standard LPC estimator is 4dB SNR, whereas the pole enhancement technique fails at -6dB SNR, representing an improvement of 10dB noise immunity in favour of the pole enhancement technique.

The performance curves for F2 show an apparent decrease in noise immunity improvement for the pole enhancement method. The standard LPC analysis ostensibly fails the difference limen of 50Hz at 5dB SNR whereas the pole enhancement technique fails at -3dB, representing an improvement of 8dB. However, the standard LPC curve is affected below 11dB SNR by a drop in the number of formants detected as representing F2, and Fig. 1(c) demonstrates the percentage of total frames for each technique in which a value for F2 was able to be extracted. Thus, using 10dB as the limiting SNR for the standard LPC technique in this case yields a noise immunity improvement of 13dB in favour of the pole enhancement technique. Results for F3, although not shown here, demonstrated a noise immunity improvement of 8dB in favour of the pole enhancement technique.

In order to explore the separation of merging formants and integrity of formant tracking, a real speech signal was used. Here, the performance of the pole enhancement technique is compared against that of an LPC-based formant tracking system which also employs off-axis spectral estimation but which uses a fixed model order[5]. In this latter technique, if it is not possible to extract more than 3 formants from the normal unit-circle LPC smooth spectrum for the current analysis frame, then the z-transform radius is decreased in steps of 0.004, down to a minimum radius of 0.88, until frequency values are found for the missing formants. The model is held constant and there is no attempt to carry out averaging of formant data. The analysis parameters for both techniques were the same as those outlined in the noise tolerance experiment above, save that analysis frames were used at 5ms intervals. Both techniques shared the same formant line tracking algorithm [5].

The vowel segment used was the vowel /i/ segmented from the word "deed" spoken by a male speaker in continuous speech. Typical results are shown in Figs. 2(a) and 2(b), which demonstrate the superior performance of the pole enhancement technique in separating F2 and F3 even in a very favorable SNR environment (here, SNR = 60dB). (The application of a wideband spectrogram fails to separate F2 and F3).

AN IMPROVED FORMANT TRACKING ALGORITHM FEATURING ADAPTIVE POLE ENHANCEMENT

Experimentation with the same speech signal contaminated with additive white noise has found that the pole enhancement technique provides an additional 14dB of noise immunity in maintaining the integrity of the F2 formant track, i.e. until the track first begins to break up.

SUMMARY AND CONCLUSION

A novel formant estimation technique has been presented which is based on linear prediction coding, but improves formant estimates by seeking to enhance the spectral effect of poles in the vocal tract transfer function by using off-axis z-transformations. The use of an increasing LPC model order as z-transform radius decreases helps separate merged formants, and post-filtering the results with a simple averaging filter enhances the effects of stationary features in the short-time speech waveform at the expense of nonstationary spectral features associated with additive broadband noise.

The increase in noise immunity afforded by the pole enhancement technique in its rôle as a spectrally-based formant estimator has been demonstrated to be in the region of approximately 10-15dB depending on the formant. This improvement in formant estimation has been found to similarly improve the noise immunity of the associated line tracking algorithm.

Lastly, the processing mechanics of the pole enhancement technique provide the formant estimation process with an inherent relative insensitivity to model order as compared to standard LPC analysis. This can be considered to make the overall performance of the formant tracking system less sensitive to variations in the acoustic characteristics of different talkers.

This work has been supported by a U.K. Science and Engineering Research Council grant.

REFERENCES

[1] J.D. Markel, "Digital inverse filtering- A new tool for formant trajectory estimation", I.E.E.E. Trans. AU, v.AU-20(2), pp.129-137 (June 1972)

[2] J. Makhoul, "Linear prediction- a tutorial review", Proceedings of the I.E.E.E., v.63, pp.561-581 (April 1975)

[3] J.L. Flanagan, "Difference limen for vowel formant frequency", Journal of the Acoustical Society of America, v.27, pp.613-617 (1955)

[4] P. Mermelstein, "Difference limens for formant frequencies for steady-state and consonant-bound vowels", Journal of the Acoustical Society of America, v.63(2), pp.572-580 (Feb. 1978)

[5] S.S. McCandless, "An algorithm for formant extraction using linear prediction spectra", I.E.E.E. Trans. ASSP, v.ASSP-22, pp.135-141 (April 1974)

AN IMPROVED FORMANT TRACKING ALGORITHM FEATURING ADAPTIVE POLE ENHANCEMENT

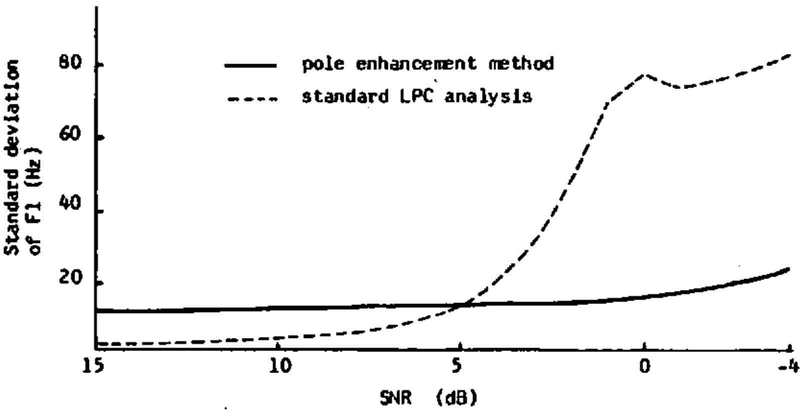

Fig. 1(a)

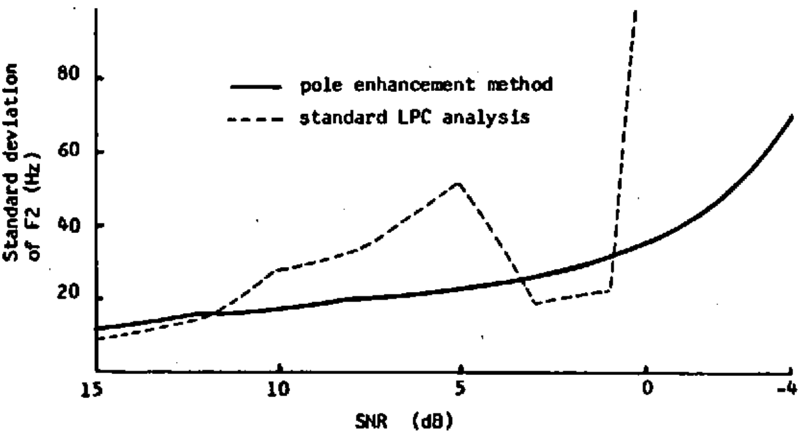

Fig. 1(b)

Over../

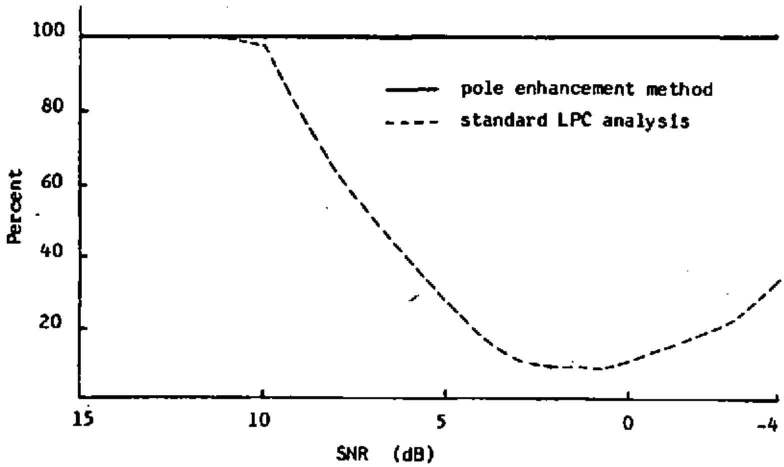

Fig. 1(c)

Fig. 1 (a-b) Standard deviation of (a) F_1 and (b) F_2 formant estimates against signal-to-noise ratio for the pole enhancement method and standard LPC analysis using a 16th order model.

(c) Percentage of analysis frames containing estimates for F_2 against signal-to-noise ratio for the pole enhancement method and standard LPC analysis using a 16th order model.

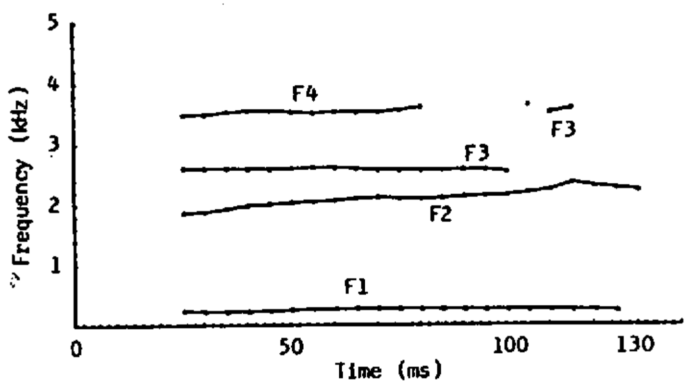

Fig. 2(a)

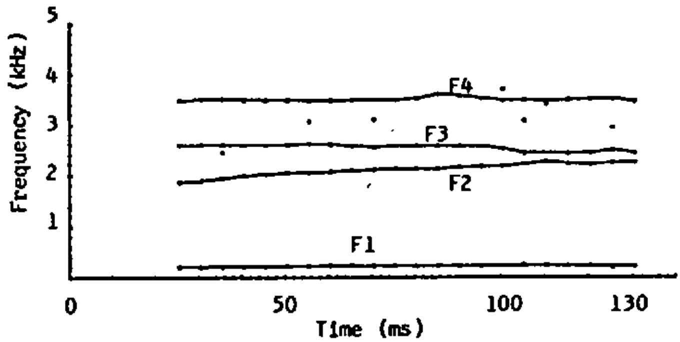

Fig. 2(b)

Fig. 2 (a) Formant tracks for F_1 - F_4 using an LPC-based spectral estimator employing off-axis spectral estimation but with a fixed 16^{th}-order model. SNR≙60dB.

(b) Formant tracks for F_1 - F_4 using the pole enhancement technique.