POLE-ZERO ANALYSIS FOR THE DETECTION OF NASALITY

C. D. Summerfield †

†School of Electrical Engineering,
The University of Sydney

ABSTRACT - The detection of nasality and the fine-class categorisation of nasal segments is important for the success of a phonetically based speech recognition machine. In this paper the application of a pole-zero modelling algorithm to this problem is described. It is well known that nasal segments are characterised and may be classified by the presence and location of the vocal tract transfer function zeros. The aim, in applying the pole-zero algorithm, is to elucidate this particular acoustic feature. However, as will be demonstrated, the zero response from the pole-zero algorithm show considerable amount of extra activity which is not attributed to the vocal tract zero alone. These results indicate that the application of the pole-zero algorithm yields results which are difficult to interpret and, consequently of limited use in this application.

INTRODUCTION

The acoustic theory of speech production predicts that the transfer characteristic of an acoustically isolated vocal tract excited at the glottis is modelled by an all-pole transfer function (Fant, 1960). This model is based on a number of assumptions about the physical configuration of the vocal tract system and the point of excitation. The most important assumption in this model in the study of nasality is the notion that the vocal tract is acoustically isolated and contains no side branch resonance chambers. During the production of nasal segments the oral tract is completely sealed and the velum is lowered to form an acoustic channel through the nasal cavity. In this configuration, the closed oral cavity constitutes an acoustic side branch to the main acoustic channel through the nasal passage. This side branching cavity absorbs energy from the main acoustic channel at frequencies close to the resonance frequency of the chamber, thus, introducing zeros into the vocal tract system transfer characteristic. The location of the transfer function zeros in the s-plane (or z-plane in discrete systems) introduced by this side branch are determined by the physical dimensions of the oral cavity during the nasal segment production. The physical length from the velic port to the point of occlusion of the shunting oral side branch cavity determines the resonance frequency of the cavity and, thereby, depicts the position of the zeros on the z-plane. Altering the position at which the oral cavity is sealed, therefore, influences the position of the zeros in the vocal tract systems transfer function. As the length of the shunting oral cavity is reduced (ie as the place of articulation is moved back) the first resonance frequency of the cavity is increased, thus, increasing the frequency of the zeros in the transfer function. Measurements of the length of the oral cavity during the production of various nasal segments indicate that the typical frequencies at which zeros are predicted to appear for the nasal consonants /m/, /n/ and /ng/ are approximately 1 kHz, 1.7 kHz and 3 kHz, respectively (Fujimmura, 1962). Although there is some variance in the results reported by other phoneticians, it is generally agreed that, as the "place of articulation" is moved back towards the velic port the higher the frequency of the zero attributed to the shunting effect of the sealed oral cavity.

The transfer function of the supra-glottal system during the production of nasal segments is, therefore, described by a transfer function containing both pole and zero components, $H_p(z)$ and $H_z(z)$ respectively,

$$H(z) = G\frac{H_z(z)}{H_p(z)}. \tag{1}$$

Where G is the gain of the system. In this model the distribution of the poles (or roots) of the denominator polynomial are determined mainly by the dimensions of the pharynx and nasal cavities which constitute the main acoustic channel, whilst the zeros (or roots) of the numerator polynomial are determined by the dimensions of the acoustic side branch, in this case the sealed oral cavity.

It should be stated that the acoustic coupling of side branch cavities is not the only source of transfer function zeros in the vocal apparatus. The progressive deviation from the plane wave model of sound propagation in the vocal tract at high frequencies leads to the introduction of significant transfer function zeros at frequencies above about 3 kHz. Major transfer function zeros are also introduced when the excitation function is moved from the glottis to some point in the vocal tract, as is the typical situation during the production of fricatives and the burst release of stops and affricates. For these reason this investigation was confined to the zero response below 3 kHz and was performed on voiced segments only.

POLE-ZERO ALGORITHM

In order to investigate the possibility of using the zero distribution to detect and classify nasality it is necessary to use an analysis algorithm which can evaluate both the pole and zero polynomials in equation 1. The pole-zero analysis method used in the present study is based on the decomposition of the pole and zero components using the properties of the negative derivative of phase spectrum (NDPS) (Yegnanarayana, 1981).

In the NDPS representation the response of the pole part of the transfer function $H_p(z)$ is almost entirely confined to the positive half of the plane, whilst the corresponding response of the zero part $(H_z(z))$ is confined to the negative half of the plane. The actual operation of decomposition is simply performed by splitting the NDPS representation into the respective positive and negative components along the NDPS=0 axis. The separate positive and negative responses are then used to compute the coefficients for the denominator and numerator polynomials, respectively.

The NDPS is derived by computing the forward Fourier Transform of a "ramp" weighted cepstral response. The "ramp" function is truncated in this algorithm to produce a smoothed NDPS and controls the modelling detail in the output pole-zero analysis. To decomposes the pole and zero components the NDPS is split along in the manner described above. The respective pole and zero cepstral responses are then determined by taking the forward transform of the respective halves of the split NDPS and applying an inverse "ramp" function weighting to cancel the original cepstral weighting function. The coefficients of the respective numerator and denominator polynomials are computed from the respective, pole and zero cepstral responses using a reverse recursion procedure. The separate pole and zero transfer function characteristics are computed directly from the denominator and numerator polynomial coefficients.

The pole-zero algorithm has been implemented in the 'C' programming language and installed in the AUDLAB interactive speech and signal processing environment (Terry et al, 1987).

RESULTS

The logarithmic zero transfer functions computed by the pole-zero algorithm for nasal segments /m/, /n/ and /ng/ produced in /iː/, /aː/ and /uː/ vowel contexts are shown in figure 1 a, b, and c, respectively. These spectral responses were produced using a 20 mS Hamming window placed at the mid point in the nasal segment. The pole-zero analysis used a 512 point NDPS and a 40 point pole-zero modelling order. The speech waveform was sampled at 16 kHz. The speech segments used in these examples were produced by an adult male RP speaker.

In an attempt to identify the salient zeros attributed to the side branching of the oral cavity a sequence of zero transfer function responses were computed across the complete nasal segment at 5 mS analysis intervals. The sequence of zero transfer function responses were analysised by a dip detection algorithm which logged the frequency location and minimum gain of the transfer function zeros over the frequency range 0 to 3 kHz. The results of the zero detection were logged using a set of contiguously frequency bins located at 100 Hz frequency intervals.

To emphasis the stronger zeros in the responses the frequency bins is incremented by a factor inversely proportional to the minimum gain of the zero. This produces a zero index which increases as the depth of the zero also increases. The variability in the duration of each segment is eliminated from the procedure by taking the average of the accumulated indices in each frequency bin. Finally the results from the analysis of each individual nasal segments in the various vowel-contexts are combined to produce an overall response for each nasal segment type. The combined results for segment types, /m/, /n/ and /ng/ are shown in figures 2a, b and c, respectively.

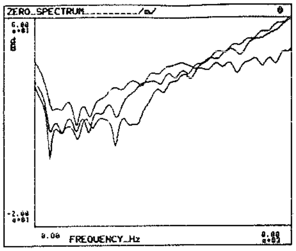

Figure 1a: Zero transfer function for the bilabial /m/

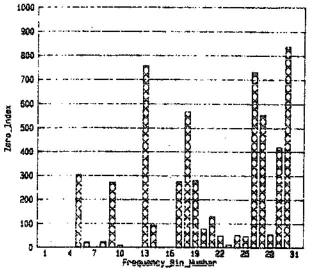

Figure 2b: Zero index for the bilabial /m/

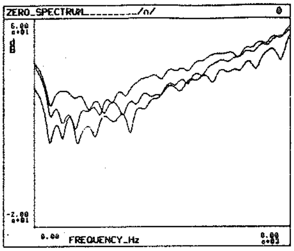

Figure 1b: Zero transfer function for the alveolar /n/

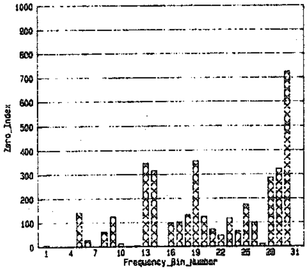

Figure 2b: Zero index for the alveolar /n/

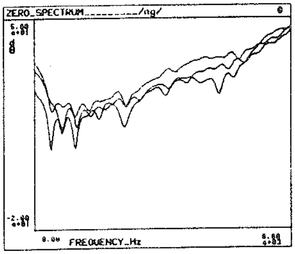

Figure 1c: Zero transfer function for the velar /ng/

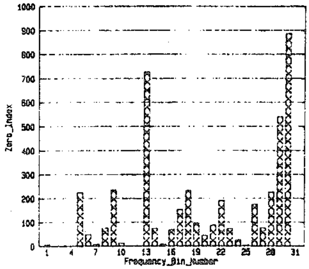

Figure 2c: Zero index for the velar /ng/

COMMENTS

The results shown in figures 1a, b and c indicate that the zero activity reflected in the zero transfer function of the pole-zero analysis is far more complex and contains far more information than can be attributed to the shunting effect of the oral cavity alone. The similarity in the results presented in figure 2a, b and c also indicate that there is little systematic difference in the distribution of the zeros upon which to base fine-class categorisation of the nasal segments. Initial results from the analysis of other subjects also tend to substantiate this finding.

Although the all-zero modelling of the speech signal is unsuccessful in detecting and categorising nasal segments, some success has been achieved in it use for determining vowel nasalization, particularly in open vowels. Analysis of open vowels produced just prior to the nasal segements consistently showed a strong zero contribution which appeared at approximately 500 Hz when compared to the same vowel in a non-nasal context. Experimental results from the all-pole modelling of the open vowel show that this zero is approximately 7dB below the same response found in non-nasalised vowels.

One of the main problems in attempting to use the zero transfer function to infer information about the supra-glottal configuration in voiced speech is that the spectral characteristics of the glottal sound source is dominated by zero contributions. Acoustic coupling of the sub-glottal system to the supra-glottal structures during the open phase of the glottal pulse. As the sub-glottal system constitutes a large cavity behind the excitation source it will, consequently, introduce spectral zeros into the transfer function model of the complete speech production system. This contribution of the glottal source function compounds the zero masking problem caused by the spectral characteristics of the glottal pulse itself.

It is not clear from these results that the zero spectrum does actually contain the information about the spectral properties of the side branch structures of the vocal system. The dominance of the zero activity attributed to the glottal source function masks the acoustic features useful for detecting nasality. It is clear that further research is required on the algorithm and its application to minimise the effects attributed to the source filter interaction before it can be fully exploited in this application.

ACKNOWLEDGEMENTS

This work was conducted at the Centre for Speech Technology Research, University of Edinburgh. I would like to acknowledge the kind help and advise of the Director of the Centre Professor John Laver and and my colleague, Dr. Steve Hiller.

REFERENCES

Fant G. (1960). *Acoustic Theory of Speech Production,* Mouton, The Hague.

Fujimmura O. (1962) "Analysis of nasal consonants", JASA 34, pp 1865-1987.

Yegnanarayana B. (1981) "Speech Analysis by Pole-zero Decomposition of Short-Time Spectra", Signal Processing 3, North-Holland.

Terry M. et al (1987) "The AUDLAB Interactive Speech Analysis System", 1st IASTED International Symposium on Signal Processing and its Applications, pages 614-617.