SPEECH PRODUCTION MODELLING WITH VARIABLE GLOTTAL REFLECTION COEFFICIENT

D M Brookes and P A Naylor

Department of Electrical Engineering
Imperial College of Science and Technology, London SW7, England.

Abstract

Recognition and synthesis of speech require accurate models of speech production. Inherent in linear predictive modelling is the assumption that the formant frequencies and bandwidths of speech do not change within the analysis frame - usually at least a larynx cycle. It is shown that there can be significant differences in formant characteristics between the closed and open glottis phases of the larynx cycle.
A development of the lossless tube model is presented in which the formant variations between closed and open glottis phases are modelled by a time varying glottal reflection coefficient. A numerically based method for estimating the parameters of such a model is outlined. The results of analysis and re-synthesis of female voiced speech obtained using this model are compared to those obtained using conventional LPC.

1. Introduction

The most commonly used models of speech production represent the vocal tract by a linear predictor [1] [5]. Such models - parametrically described by predictor coefficients - can give excellent results in speech coding applications. However, three simplifying assumptions, normally inherent in linear predictive models of the vocal tract, may lead to model inaccuracies which are more serious in recognition and synthesis applications than in coding.

(a) Linear predictive analysis assumes that the vocal tract can be represented by an all-pole filter.

(b) In the simplest case, linear predictive analysis assumes impulse excitation of the vocal tract.

(c) It is assumed in linear predictive models that the vocal tract transfer function is constant for at least a larynx cycle. However, observation and theoretical studies indicate that the frequencies and bandwidths of formants change within a larynx cycle. This effect can be seen in the example of figure 1 obtained using covariance LPC on the closed glottis and open glottis segments of the speech wave.

Many improvements have been made to the simple LPC model in which assumptions (a) [9] or (b) [6] are not made. In the model described in this paper assumptions (a) and (b) are retained, but assumption (c) is no longer made.

Closed phase analysis has been proposed as an improvement to standard linear predictive analysis [10][4]. In this case, the coefficient estimates are calculated only while the glottis is closed and excitation is zero. These estimates are then assumed valid for the whole cycle, including the open glottis phase. The use of closed phase analysis can result in better estimates of the vocal tract transfer function and in more consistent formant tracks for two reasons. Firstly, the analysis time is short enough that the formant frequencies and bandwidths are almost constant, and secondly, the vocal tract filter estimate is not affected by the spectral content of the excitation. The formant estimates for the open glottis phase are inevitably wrong, but modelling accuracy in the open phase is less important because the speech amplitude is normally low.

For some speakers, the true closed phase may be extremely short or even be entirely absent [2]. In these cases insufficient data is available for the reliable estimation of closed phase predictor coefficients. This problem arises particularly with female speech when the fundamental frequency is high.

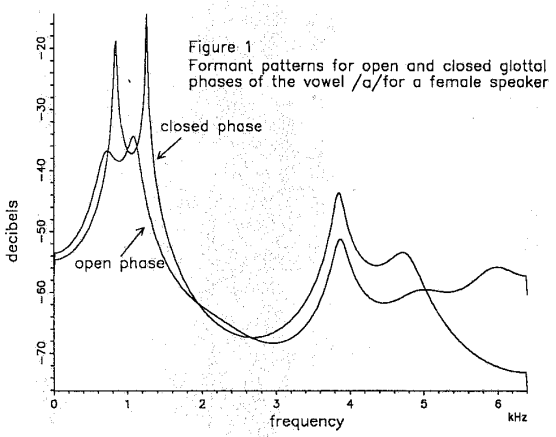


Figure 1
Formant patterns for open and closed glottal phases of the vowel /a/for a female speaker

CH2561-9/88/0000-0671 $1.00 © 1988 IEEE

This paper presents a development of the lossless tube model of speech production in which the changes of formant frequencies and bandwidths that occur within a larynx cycle are modelled by a variable glottal reflection coefficient. The technique presented uses all the data in each cycle for parameter estimation and does not, therefore, suffer from the above problems that arise with closed phase analysis.

2. The Lossless Tube Model

The vocal tract can be modelled by a concatenation of p lossless tubes as shown in the 4 pole example of figure 2. Changes in cross-sectional area at the tube boundaries can be represented by p+1 reflection coefficients. If the volume velocity of air flow at the glottis is Ug and at the lips is U_l then the transfer function can be written [8]

$$\frac{U_l}{U_g} = \frac{0.5(1 + r[g]) \prod_{k=1}^{p} (1 + r[k])\, z^{-p/2}}{D[z]} \qquad (1)$$

where

$$D[z] = \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_p \\ -r_p z^{-1} & -z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad (2)$$

It can be shown that D[z] is of the form

$$D[z] = 1 - \sum_{k=1}^{p} a[k] z^{-k} \qquad (3)$$

where the a[k] are the LPC predictor coefficients. For example, when p = 3,

$$D[z] = 1 - (r[1]r[2] + r[2]r[3] + r[1]r[g])\, z^{-1} - (r[1]r[3] + r[1]r[2]r[3]r[g] + r[2]r[g])\, z^{-2} - (r[3]r[g])\, z^{-3} \qquad (4)$$

Conventional LPC analysis determines the predictor coefficients a[k] which give minimum mean square prediction error. If a value for r[g] is known or assumed then the remaining reflection coefficients may be obtained from the a[k] by equating the polynomials in (3) and (4). For the particular case of r[g] = 1, a simple recursion formula can be used.

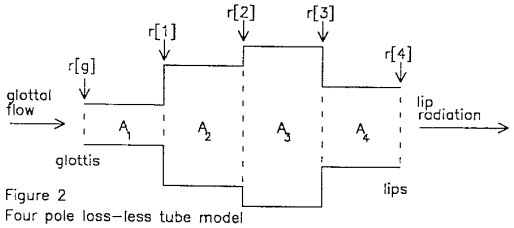


Figure 2
Four pole loss-less tube model with cross-sectional areas A_k and reflection coefficients r[i]

Theoretical studies attribute the effects seen in figure 1 to absorption of energy in the large sub-glottal volume during the open phase of the larynx cycle. This effect can be represented in the lossless tube model by allowing the glottal reflection coefficient to vary within the larynx cycle.

In the model described here, the glottal reflection coefficient r[g] is assumed to be real and is allowed to take two values - one in the closed glottis phase and one in the open glottis phase. The remaining coefficients are held constant throughout each larynx cycle. Two forms of excitation have been used. In the first form, a single impulse is applied at the instant of glottal closure. In the second form two impulses are used: one at closure and one at opening. The times of glottal closure (tc) and opening (to) are determined initially from a laryngograph waveform (Lx or EGG) [3]. The analysis procedure is then required to determine the p+2 reflection coefficients, the excitation impulse amplitudes (gc, go) and, if required, the closure and opening times (tc, to) in order to minimise the mean square re-synthesis error.

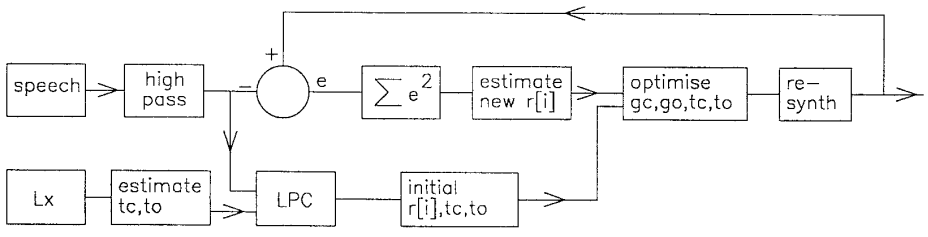


Figure 3 Optimum parameter estimation method for the loss-less tube model

3. Parameter estimation method

The model described above imposes the constraint that all but one of the reflection coefficients remain constant throughout the entire larynx cycle. From equations (3) and (4) it can be seen that this imposes a set of non-linear relationships between the predictor coefficients, a[k], in the closed phase and those in the open phase. It is not therefore possible to determine the a[k] by solving a set of linear equations as is done in LPC analysis.

An iterative method, illustrated in figure 3, has been developed to determine the r[i] directly from the speech waveform. Because this problem is intrinsically non-linear, no computational penalty is incurred in minimising the actual re-synthesis error rather than, as is done in LPC, the prediction error.

For each cycle, initial values for the reflection coefficients are determined from the results of conventional LPC analysis [8]. Reflection coefficient sets estimated in this way always yield r[g] = 1.0 for the closed phase and no value for r[g] in the open phase. An initial estimate of the latter parameter can be made arbitrarily (say 0.5) or by using the value obtained for the previous cycle. In each iteration, a new set of reflection coefficients is estimated and, for this set, the values of gc, go, and, if required, tc and to are calculated to minimise the mean square re-synthesis error. The error calculated is used in the generation of the reflection coefficient estimates for the next iteration [7] and the process is repeated until the global error minimum is found.

4. Results

The analysis and re-synthesis procedure described in section 3 has been applied to the vowel /ɑ/ from female speech and the results for three cycles are presented. Larynx synchronous LPC with impulse excitation is presented for comparison.

(a) natural speech

(b) re-synthesised speech, 16 pole LPC with
- single impulse excitation
- tc determined from Lx

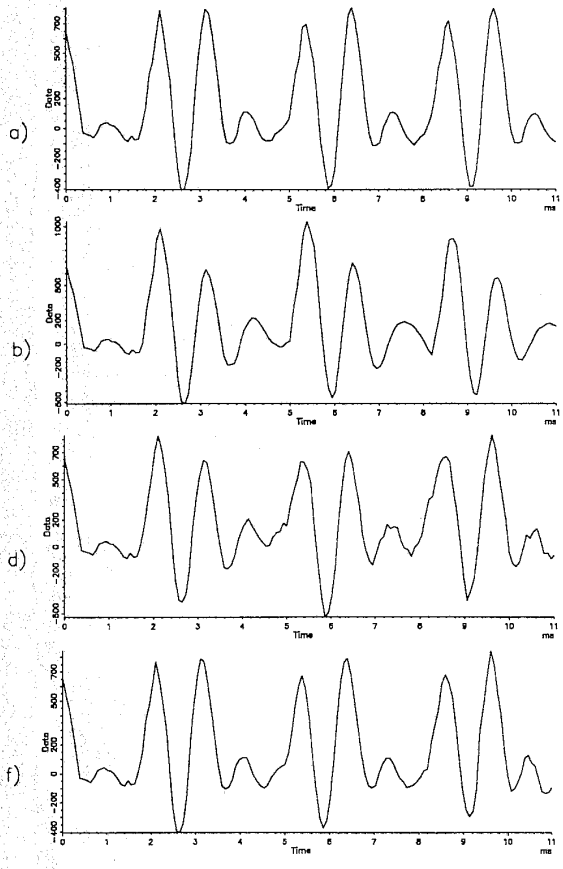
(c) re-synthesis error: (a) - (b)

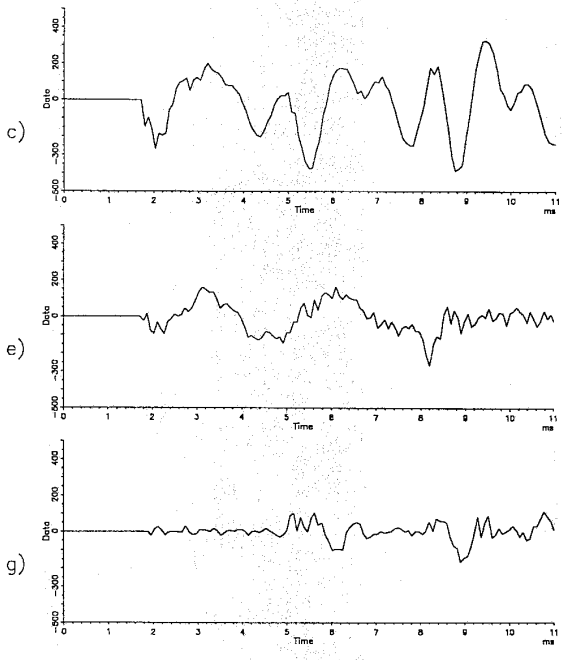
(d) re-synthesised speech, 14 pole model with
- variable r[g]
- single impulse excitation
- tc and to determined from Lx

(e) re-synthesis error: (a) - (d)

(f) re-synthesised speech, 14 pole model with
- variable r[g]
- dual impulse excitation
- optimised tc, to

(g) re-synthesis error: (a) - (f)





The RMS values of the original speech waveform and of the three error waveforms (in consistent units) are as follows:

(a) - 327
(c) - 164
(e) - 81
(g) - 48

5. Conclusions

A speech production model has been developed in which the glottal refection coefficient is varied within each larynx cycle. The model parameters are determined for minimum re-synthesis error rather than, as with LPC, the prediction error. Analysis and re-synthesis of speech using this model results in significantly lower re-synthesis error than is obtained using whole cycle, larynx synchronous LPC. The re-synthesis error can be reduced still further by optimising the times of glottal opening and closure.

Further work on the model is directed towards the inclusion of better excitation models of the excitation waveform and of the time variation of the glottal reflection coefficient.

References

[1] Atal, B. S. and Hanauer, L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am., Vol. 50, pp 637-655, Aug. 1971.

[2] Cranen, B. and Boves, C., " A Parametrical Voice Source Model Incorporating Inter and Intra- speaker Variation," Proc. IEE Int. Conf. on Speech Input/Output, Techniques and Applications, London 1986.

[3] Fourcin, A. J., "Laryngographic Assessment of Phonatary Function," ASHA Reports 11, The American Speech-Language-Hearing Association, Rockville, Maryland, 116-127.

[4] Larar, J. N., Alsaka, Y. A., Childers, D. G., "Variability in Closed Phase Analysis of Speech," Proc. of ICASSP 1985, Vol. 3, pp. 29.2.1 - 29.2.4.

[5] Makhoul, J., "Spectral Linear Prediction: Properties and Applications," IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-23, No. 3, June 1975.

[6] Milenkovic, P., "Glottal Inverse Filtering by Joint Estimation of an AR System with a Linear Input Model," IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-34, No. 1, Feb. 1986

[7] Numerical Algorithms Group, "Minimizing or Maximizing a Function," NAG Library Manual, Mark 11, Vol. 3, routine E04JAF, 1986.

[8] Rabiner, L. R. and Schafer, R. W., "Digital Processing of Speech Signals," chapts. 3 and 8, Prentice-Hall, Englewood Cliffs, N.J., 1978.

[9] Song, K. H., and Un, C. K., "Pole-Zero Modelling of Speech Based on High-Order Pole Model Fitting and Decomposition Method," IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-31, pp 1556-1565, Dec. 1983.

[10] Whitaker, L. C., and Pearce, D. J. B.,"Larynx Synchronous Formant Analysis," Proc. European Conf. on Speech Technology, Edinburgh, Sept. 1987

The work presented in this paper forms part of the SPAR *research programme.*