Pole-Zero Modeling of Vocal Tract for Fricative Sounds

Minsheng Liu, Arild Lacroix
Institut fuer Angewandte Physik, University of Frankfurt
Robert-Mayer-Str.2-4; 60325 Frankfurt am Main, Germany
e-mail: Liu@iap.uni-frankfurt.de,
Lacroix@iap.uni-frankfurt.de

ABSTRACT

This paper presents a pole-zero model based on a multi-tube acoustic model for fricative sounds. This model consists of the front and back cavity formed by oral tract and pharynx, in which the excitation source is located at the point of constriction. The transfer function of this model including poles and zeros is derived and its properties are investigated. Small losses such as viscous friction which is an important for the fricative sound in the vocal tract are considered and the results show, if the vocal tract is lossless, the numerator part of the pole-zero model is symmetric. The transfer function with small losses overcomes the limitation of the symmetry. This method is applied by employing the inverse filtering and an adaptive algorithm to analyse fricative sounds.

1 INTRODUCTION

For fricative sounds the excitation source is inside the vocal tract, where the acoustic waves propagate in two directions. Thus the vocal tract is separated by the source of excitation at the constriction into two cavities. The back cavity traps energy and introduces antiresonances, resulting in zeros in the transfer function. In this case the LPC method widely used in the processing of speech signals is not properly adapted.
The method of acoustic modeling has its advantages [1] and in [3] a simple model for fricative sounds is discussed. The source location for fricative sounds has been investigated [2]. In [4] an improved vocal tract model is used to represent nasal sounds. Till now transfer functions characterized by anti-resonances as well as resonances for fricative sounds have not been derived and computed on the basis of the acoustic tube model. In this contribution the acoustic tube model is improved by using a three-port adaptor at the position of the excitation, therefore a proper transfer function including zeros in the frequency response is derived and zeros can be calculated from the numerator polynomial of the transfer function.

2 POLE-ZERO MODELING FOR FRICATIVE SOUNDS

The unvoiced fricatives [f], [s] and [ʃ] are produced by a steady air flow which becomes turbulent in the region of a constriction within the vocal tract. The location of the constriction determines which fricative sound is produced. For the fricative [f] the constriction is near the lips, for [s] it is near the teeth; and for [ʃ] it is near the back of the oral tract. Sounds are radiated from the lips. In this case the vocal tract can be characterized by the acoustic tube model depicted in fig. 1. The excitation source is located at the junction of the section between two tubes.

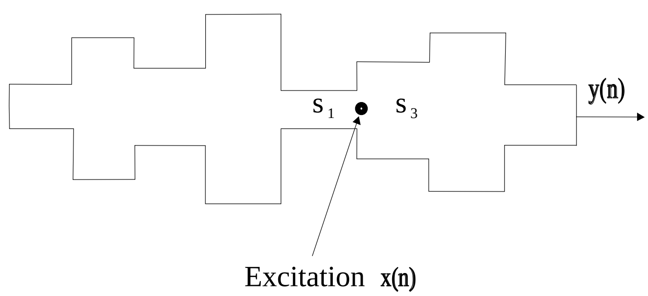

Fig. 1 Acoustic tube model for production of fricative.

We can see that the acoustic wave at the point of excitation propagates forward into the front cavity and backward into the back cavity. Together with the excitation source, there are three wave-directions at the location of the excitation source.Thus we use a three-port-adaptor to describe the acoustic characteristics at the excitation source, as shown in fig. 2. The relationship between the traveling waves in the remaining adjacent tubes can be represented by two-port-adaptors[5].

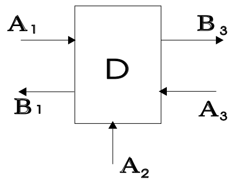

Fig. 2 Three-port adaptor at excitation source

The entire vocal tract can be realized using two- and three-port-adaptor in discrete-time domain as shown in fig. 3.

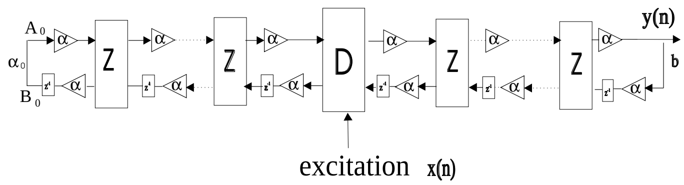

Fig. 3 Digital realization of the model in Fig. 1 in discrete-time domain

the constants α are lossy coefficients; Z and D express the two- and three-port-adaptor; the constant α_0 describes the opening of the glottis, its value lies in [-1,1].

Derivation of the Transfer Function

The three-port adaptor can be presented by a scattering matrix from the continuity equations for pressure and flow [6]. The linear relation for the flow is given by:

$$\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} = \mathbf{S} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \tag{1}$$

with

$$\mathbf{S} = \begin{pmatrix} \alpha_1 - 1 & \alpha_1 & \alpha_1 \\ \alpha_2 & \alpha_2 - 1 & \alpha_2 \\ \alpha_3 & \alpha_3 & \alpha_3 - 1 \end{pmatrix} \tag{2}$$

and

$$\alpha_i = \frac{S_i}{S_1 + S_2 + S_3} \tag{3}$$

S_1 and S_3 are the cross sectional areas of the front and the back cavity and S_2 describes the coupling of the excitation source to the vocal tract.

The acoustic wave at the end of the back cavity is partly reflected; there is the relation $A_0 = \alpha_0 \; B_0$ and the equation(1) can be transformed into the following equation:

$$\begin{pmatrix} B_2 \\ A_2 \end{pmatrix} = \mathbf{T_D} \begin{pmatrix} A_3 \\ B_3 \end{pmatrix} \tag{4}$$

$\mathbf{T_D}$ is a 2 x 2 matrix and is the function of the scattering transfer matrix of the back cavity, its elements are fractions.

The transfer function is finally derived and can be expressed as:

$$H(z) = \frac{1}{(0,1)\mathbf{T_D}\mathbf{T_F}\begin{pmatrix} -1 \\ 1 \end{pmatrix}} \tag{5}$$

$\mathbf{T_F}$ is the scattering transfer matrix of the front cavity.

If we have chosen the back cavity consisting of two sectional tubes, the first tube is a single section tube, the number of sections of the second tube is n; then the transfer function can be expressed:

$$H(z) = \frac{\alpha_0 z^{n+1} + \alpha_0 \alpha_1^2 r_0 z^n + \alpha_2^2 r_0 z + \alpha_1^2 \alpha_2^2}{(0,1)\mathbf{T_{DD}}\mathbf{T_F}\begin{pmatrix} -1 \\ 1 \end{pmatrix}} \tag{6}$$

r_0 is the reflection coefficient between the two tubes in the back cavity; α_1 and α_2 are the lossy coefficients of the back cavity. $\mathbf{T_{DD}}$ is a 2x2 matrix and its elements are polynomials.

Properties of the Transfer Function

We can see that the numerator of the transfer function is a polynomial containing the zeros in the transfer function. From this polynomial we can calculate the zero locations after estimating the model parameters.

It can be seen that the back cavity determines the zero locations of the transfer function, while the front cavity has no effect on it.

It can be seen that if the vocal tract is lossless($\alpha_1 = 1, \alpha_2 = 1$) and the glottis is closed or open($\alpha_0 = 1, \alpha_0 = -1$), the numerator polynomial of the transfer function is symmetric or antisymmetric,

$$z^{n+1} + r_0 z^n \pm r_0 z \pm = 0 \tag{7}$$

To break the symmetric or antisymmtric relation there are two ways, either a loss is to take into account in vocal tract or the factor α_0 of the glottis changes in (-1, 1).

If the area at the junction is $S_1 = 0$, the back cavity is eliminated, and the matrix $\mathbf{T_D}$ becomes a standard scattering transfer matrix of a two-port adaptor and the pole-zero model is transformed into the well-known all-pole model.

3 EXPERIMENTAL RESULTS AND CONCLUSION

The total lengths of the pole-zero models are chosen according to the fricative sounds. The block length for the analysis of speech signals is chosen as 20 ms coressponding to 160 sampling values at a sampling frequency of 8 kHz.

The pole-zero model is inverted and we can get the inverse structure of the filter, as displayed in fig. 4.

The model parameters are analyzed and optimized from natural speech. We have used the inverse filtering method [7] to calculate the parameters of the front cavity and an adaptive algorithm to optimize the parameters for the excitation source and the back cavity.

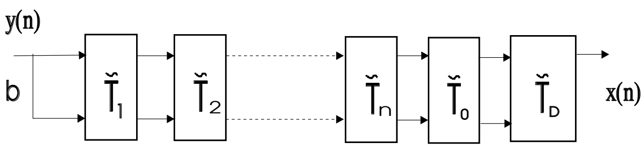

Fig. 4 Inverse filtering structure of the filter in Fig. 3

First of all we generate a synthetic signal using a prescribed pole-zero model, then we determine the reflection coefficients from the synthetic signal. A comparison of the original and calculated frequency response is shown in fig. 5. It is shown that this model has five resonances and the zeros appear in the frequency response of the original and the calculated model. Both zeros of the analytic result compared with the amplitude of the original frequency response are shifted torwards higher frequencies.

Then this pole-zero model is applied to investigate the fricatives. The speech material is produced by a male speaker. The results of two fricatives are presented in this paper.

The first example is the frictive [f]; its model is given by a total of ten tubes with the front cavity consisting of 2 tubes and the back cavity consisting of 8 tubes of different cross sections. The excitation source is located between the second and the third tube. Fig. 6 shows the analytic results. The spectrum of the speech is in good coincidence with the amplitude of the frequency response shown in fig. 6 A. The zeros that can be clearly seen in the groupdelay fig. 6 B appear also in the magnitude response.

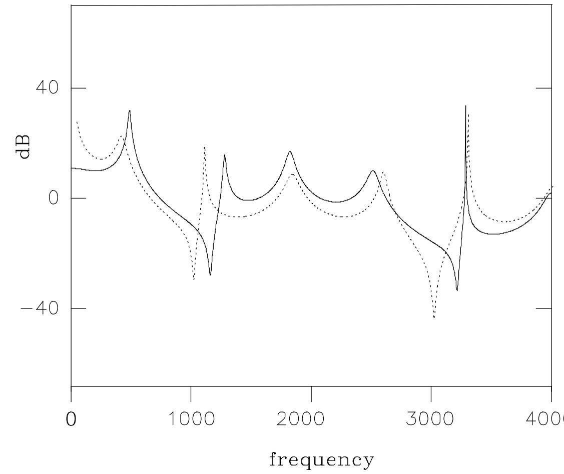

Fig.5 Comparison of the amplitude of original (solid) and analytic(dotted) frequency response.

The widths of the antiresonances and resonances are increased, if small losses are included in the vocal tract, but their locations are almost unchanged.

The second example is the fricative [$\int$]; its model is composed of 8 tubes in the front cavity and 2 tubes in the back cavity. The first tube in the back cavity is a single section tube, while the second tube consists of three tubes with same cross sections. It can be seen from in fig. 7 that the fricative [$\int$] has five resonances and two antiresonances.

The calculated results compared to the spectra indicate that the pole-zero model can properly handle fricative sounds. It turns out that the glottis in terms of α_0 is important for the adapation of the model. In future work we will investigate this effect and how α_0 can be calculated directly from speech signals.

References

[1] Schroeter, J. and Sondhi, M. M.: Techniques for Estimating Vocal-Tract Shapes from the Speech Signal, IEEE Trans. Speech and Audio Processing, Vol. 2, No. 1, pp133-150, 1994.

[2] Narayanan, S. and Alwan, A: Parameteric Hybrid Source Models for Voiced and Voiceless Fricative Consonants; Proc., ICASSP 96, pp. 377-380, May 1996.

[3] Flanagan, J. L.: Speech Analysis, Synthesis and Perception, 2nd ed., Springer-Verlag, New York, 1972.

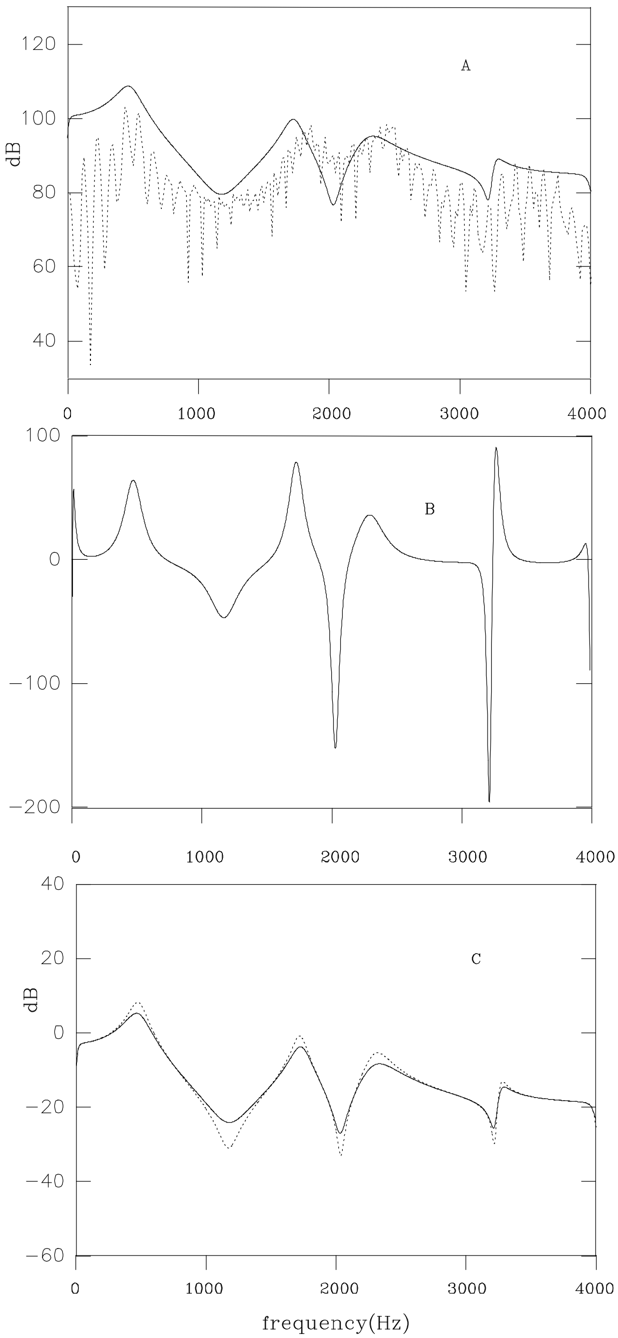

Fig.6 Analytic Results of the fricative [f]: A: spectrum(dotted line) and analytic frequency response(solid line) with small losses; B: correspondung groupdelay; C: Comparison of frequency response with small losses(solid line) and lossless(dotted line).

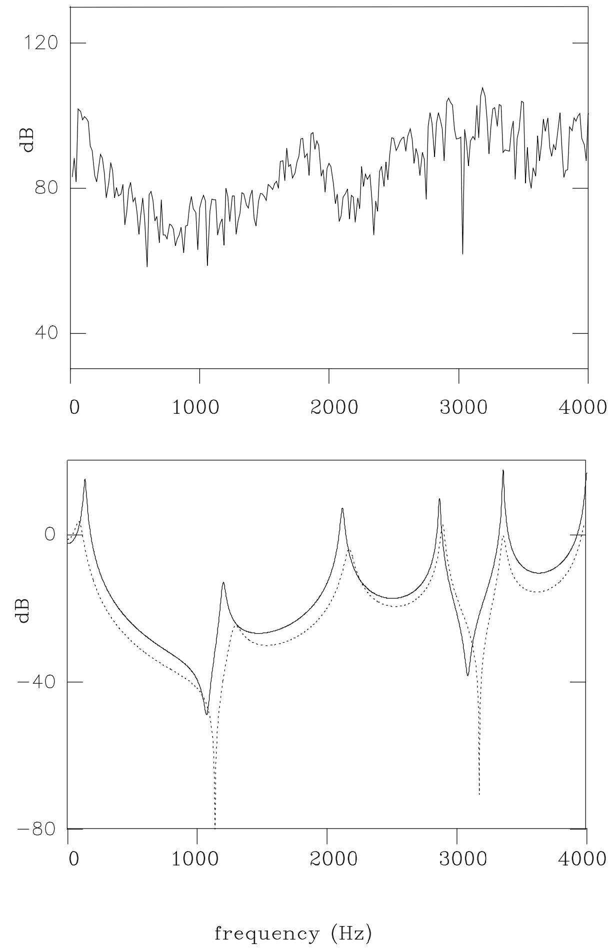

Fig.7 Comparison of spectrum (top) and analytic amplitude response (bottom) for fricative [ʃ]. Solid line: small losses; dotted line: lossless case.

[4] Liu, M., Lacroix, A.: Improved Vocal Tract Model for the Analysis of Nasal Speech Sounds, Proc. Conf., ICASSP 96, Volume:2, pp.801-804., May 1996.

[5] Lacroix, A.: Digitale Filter, 4. Auflage, R. Oldenbourg Verlag GmbH, München, 1996, pp.104-110.

[6] Hess, W.: Digitale Filter, B. G. Teubner Stuttgart 1993, pp367-401.

[7] Lacroix, A.: Ein neues Verfahren für die inverse Filterung mit dem Kreuzglied-Kettenfilter. NTZ Archiv 7, pp155-158, 1985.