

Modeling the Transitory Behavior of Speech Using a Time-varying Transmission-line Model

Amit S. Rane, Derrick C. Wei, Lisa E. Falkson, and Abeer A. Alwan

*Speech Processing and Auditory Perception Laboratory
Department of Electrical Engineering, University of California Los Angeles, Los Angeles, CA 90095.*

Abstract: In this study, a transmission-line model of speech production (1) is modified so that changes in the vocal-tract (VT) area functions, reflecting changes in the VT shape from one sound to another, are properly modeled. Each VT shape is divided into uniform cylindrical sections and each section is modeled by its equivalent transmission line comprising Resistors (R), Capacitors (C), and Inductors (L). A radiation impedance terminates the transmission-line sections. For voiced sounds, the KLSYN88 glottal waveform (2) is used as input. The model is simulated in time and frequency by using the analog-circuit simulator HSPICE. To simulate the transitory characteristics of speech, different interpolations such as linear, cubic and Fourier Transform are used for area transitions. The interpolated values result in a 'sampled' area function and the elements of each section are varied according to these time-sampled values. These variations are accomplished by utilizing time-varying independent voltage sources. Examples using American English diphthongs will be presented. Formant frequencies (both steady-state and transitions) of the synthesized and natural diphthongs are in close agreement.

INTRODUCTION

The vocal tract can be modeled as a concatenation of uniform cylindrical-tube sections where each section is approximated as a lumped circuit. Figure 1 shows a lumped circuit approximation for one section of the vocal tract; the values of the components (R, L, C, and G) are from (3). This approximation is valid as long as the cross dimensions are small when compared to the wavelength of the sound produced. For most configurations, the approximation is valid up to about 4-5 kHz.

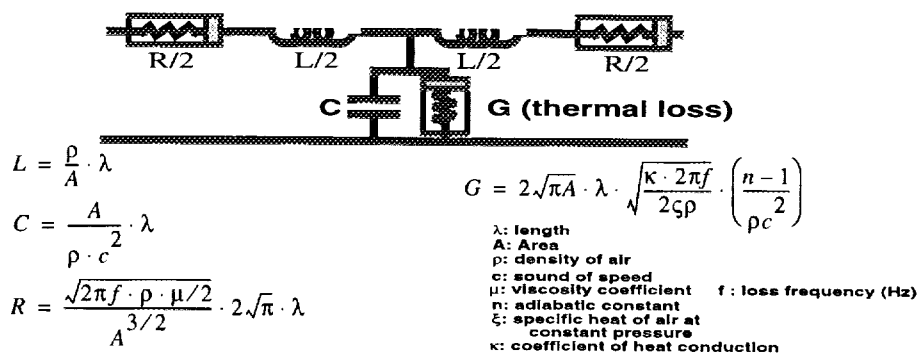


FIGURE 1. A section of the transmission line model

For voiced sounds, the transmission line model is excited by a periodic current waveform which resembles the shape of the KLSYN88 glottal waveform (2). The transmission line is terminated at the lips by a radiation impedance model proposed by Flanagan (3).

THE TIME VARYING TRANSMISSION LINE MODEL

Time varying speech sounds are produced by temporal changes in the vocal-tract shape. By time-varying the circuit elements of each section of the transmission line model, it becomes possible to capture the dynamics of the vocal tract. In the present study, the transitory behavior of the vocal tract between two vowels (*/a/* and */i/*) is analyzed, using the area functions from (4). The area of each section is varied from an initial value to a final value; these values depend upon the initial and final vowels. We chose to model the vocal tract with 35 sections and each section can be varied temporally from its initial to final value in a unique way. The area of the sections is varied by either using linear, cubic or Fourier transform interpolation. The area transition so obtained is sampled

temporally (in a uniform fashion at a rate of 16 KHz) to give a set of 'sampled areas'. The component values of each section are then varied using these 'sampled areas'. The simulation is carried out using the circuit simulator HSPICE. In order to accomplish the temporal variation, each component (L, C, R and G) in the model is made voltage variable. By varying the control voltages according to the areas, it becomes possible to simulate the effect of temporal changes in the vocal tract. Figure 2 shows the time-varying transmission line model.

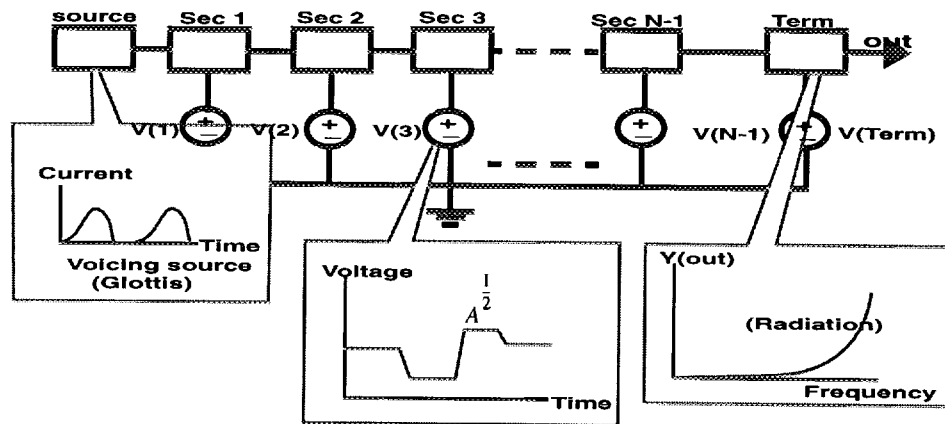


FIGURE 2. The Time Varying Transmission Line Model

RESULTS

Perceptually, linear and cubic interpolations produced good results. Due to the 'ringing' behavior of the Fourier Transform interpolation, the sounds produced due to this interpolation are not acceptable.

In the frequency domain, cubic interpolations result in the smoothest transition from /a/ to /i/ and the Fourier transform interpolation shows ringing as seen in Figure 3. Future work will investigate the synthesis of larger units, such as words, and different kinds of interpolation functions.

Work supported in part by NSF.

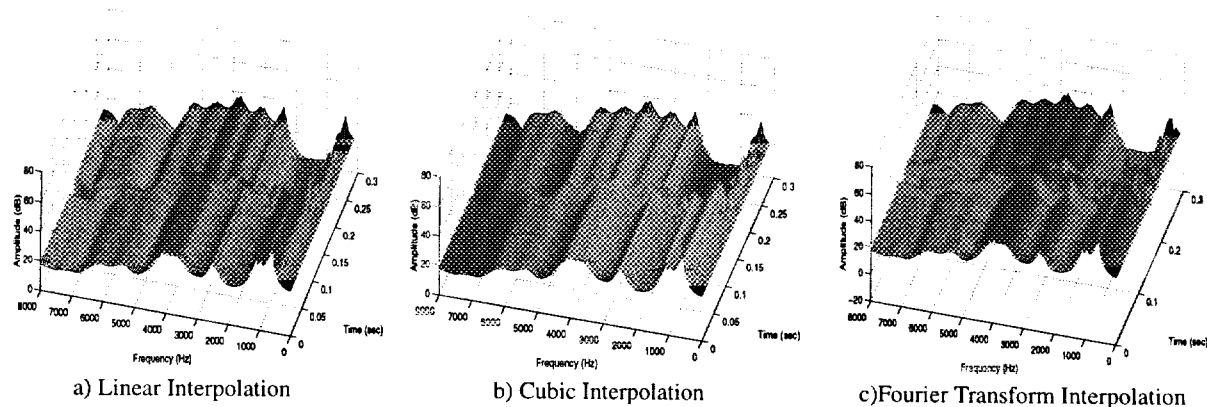


FIGURE 3. 3D Spectrograms of the Diphthong /ai/ for Different Area Interpolations

REFERENCES

1. Rael, J.J., Chang, J.Y.C. and Alwan, A., *JASA*, Supp.1, Vol. 97, (5), (1995)
2. Klatt, D.H. and Klatt, L.C., *JASA*, Vol. 87, (2), 820-857 (1990)
3. Flanagan, J.L., *Speech Analysis, Synthesis and Perception*, New York, Springer-Verlag, 1972, Chapter III.
4. Fant, G., *Acoustic Theory of Speech Production*, The Netherlands, Mouton, 1970, Chapter 2.3, pp. 115.