RECOVERING TUBE KINEMATICS USING TIME-VARYING ACOUSTIC INFORMATION

R. S. McGowan

Haskins Laboratories, New Haven, Connecticut

ABSTRACT

Formant frequency trajectories are used to optimally fit the kinematics of a modified twin-tube. An entire articulatory trajectory is fit in a single optimization, because an articulatory trajectory is modeled as a parameterized function of time.

1. INTRODUCTION

The inverse mapping between acoustics and articulation has received considerable attention in the last twenty-five years. The focus has been on mapping static spectral variables onto static vocal tract shapes, with resulting ambiguity in the mapping. Ambiguities were noted in the work of Atal, Chang, Mathews, and Tukey [1] where the articulatory positions of the vocal tract model were varied to fit formant frequency data; in the work of Flanagan, Ishizaka, and Shipley [3] using an optimization procedure based on spectral information and cepstral matching to find vocal tract area functions, as well as subglottal pressure and laryngeal parameters; and the work of Levinson and Schmidt [5] using a gradient search optimization to relate articulator positions to LPC envelopes.

Two ways of overcoming inverse mapping ambiguities suggest themselves: either decrease the number of articulatory degrees of freedom, or increase the amount of acoustic data. One procedure to decrease the number of articulatory degrees of freedom took account of the continuity of vocal tract tube shapes in short time intervals [6,4,8]. This seemed to help relieve ambiguity, but the optimizations were performed at each time sample, making the inclusion of the continuity constraint inefficient. In the method examined here, the kinematics of the articulators were parameterized as functions of time, and the optimization was performed over time spans corresponding to a single parameterization, thus the continuity constraints were automatically incorporated. Because the time spans were longer than a single time sample, there was a span of acoustic data that was used in the optimization, thus the number of degrees of freedom in the data was also increased.

2. METHOD

The acoustic data consisted of up to three formant frequency trajectories that were generated using a modified twin-tube model [2]. In the modification considered here, a third tube, a constriction tube, was placed between front and rear tubes of the twin-tube model (Fig. 1). There were five articulatory variables: front tube area, constriction tube area, rear tube area, rear tube length, and constriction tube length. The front tube length was determined by the restriction that the total tube length be 17 cm. The constriction tube could change area through time, thus opening and closing the tube between the front and rear tubes. The constriction area was parameterized as an exponential function of time. The maximum area of the constriction was assumed to be the average of the front and rear tube areas, and the minimum was zero, corresponding to complete constriction. As a result there were five articulatory kinematic parameters: the four constant articulatory variables, and the exponential growth factor for the change in constriction area (Fig. 2).

The modified twin-tube model was used for both the synthesis of formant frequency data and as a model vocal tract for articulatory kinematic parameter recovery. The relationship between the acoustic variables and the articulatory variables was given by the model function. This function was written as an implicit relation between the formant (resonance) frequencies and the articulatory variables. Thus, if the constriction area was given a trajectory, either opening or closing, it is possible to compute the corresponding formant trajectories using numerical root-solving techniques.
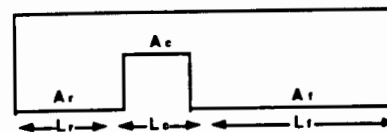

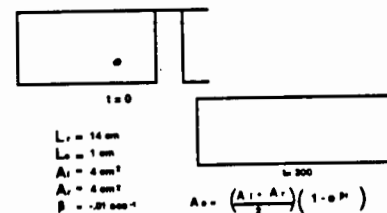Lr + Lc + Lf = 17 cm
Figure 1: Modified Twin Tube

Figure 2: Articulatory Kinematic Parameters

Preliminary work has been done on recovering articulatory kinematic parameters from synthesized formant frequency trajectories using the modified twin-tube model using a least-squares criterion. The iterative least squares was performed using the simplex method [7]. The simplex method was a conservative choice because it did not require numerical computation of a generalized inverse, as, say, the Levenberg-Marquardt algorithm did, thus reducing the possibility of numerical instability in this initial study. However, the simplex method was very slow and could be replaced with more sophisticated optimization algorithms. When the experimenter executed the program written for inverse mapping he was asked to specify the constriction length and was given the option of specifying either the front or rear tube areas. If neither of these was specified, then the optimization was performed to find four parameters: the front and rear tube areas, the rear tube length, and the exponential time constant. If one of the areas was specified, then the optimization was performed on three parameters, and if both areas were specified, then two parameters entered into the optimization: rear tube length and the exponential time constant. Because the optimization procedure was an iterative procedure that could be trapped in local minima, the simplex method was run from several initial starting places in the articulatory kinematic parameter space. The search from any of these initial starting places would terminate if the cost function was less than a given tolerance, if there was little relative change in the value of the cost function from one step to another, or if a maximum number iterations was attained.

The ideal cost function was the sum of squares of the differences over time in each formant frequency between those given by the data and the values that would be produced by the modified twin-tube model given the articulatory kinematic parameters. To have found the value of this cost function at every iteration, many formant frequencies, corresponding to a given set of articulatory kinematic trajectories, would have had to have been found. This would have involved applying root-solving techniques to the model function many times (40 times for each formant at a rate of 200 Hz for 200 msec). Accordingly, the sum of the squares of the model function evaluated at each data formant frequency was used as an alternative cost function. This appeared reasonable because it is a necessary condition that this function, being an implicit relation between formant frequency and articulatory variables, be identically zero, if the original cost function is zero.

3. RESULTS & CONCLUSION

In the modified twin-tube model, the feasibility of fitting rear tube length and exponential time constant was tested using the first formant frequency trajectory only, as well as with three formant trajectories. The feasibility of fitting four parameters, the rear tube area, front tube area, rear tube length, and exponential time constant using one and three formant frequency trajectories was also tested. As one would expect, the method did better in fitting two parameters than it did in fitting four

parameters. A counter-intuitive result is that the method seemed to have worked better with one formant (e.g. Fig. 3) than it did with three(not shown), or with less information than more. (The program was completely unsuccessful at fitting four parameters given three formant frequencies.)

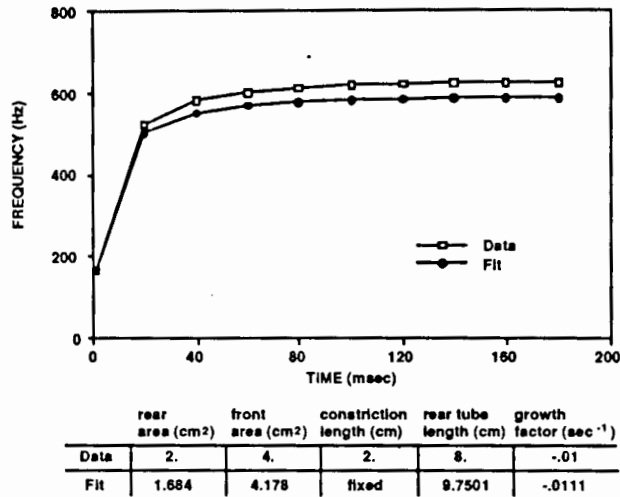|	rear area (cm²)	front area (cm²)	constriction length (cm)	rear tube length (cm)	growth factor (sec⁻¹)
Data	2.	4.	2.	8.	-.01
Fit	1.684	4.178	fixed	9.7501	-.0111

Figure 3: One resonance frequency trajectory, implicit function minimization

It was felt that something of the original cost function involving the squares of the differences between formant frequency data and those which would be produced with a given set of articulatory kinematic parameters had to be preserved to get better results. Instead of root-solving for all the formant frequency values corresponding to a given set of articulatory kinematic parameters, root-solving was performed only at the beginning, middle, and end of a trajectory for each iteration of the least-squares procedure. (For example, there were nine root solves for three formants.) The sum of squares of the differences between these frequency values and their corresponding data points were added to the sum of squares of the model function evaluated at all the data points to form a hybrid cost function. This seemed to have alleviated the counter-intuitive result of doing more poorly with three formants (Fig. 5) than with one (Fig. 4). Also, it was possible to fit the four parameters using three formant trajectories (Fig. 5).

The problem with using just the sums of squares of the model function in the cost function was that local minima ap-peared that were not close to the articulatory kinematic parameters that produced the data. By adding some explicit information to the cost function these superfluous minima no longer hindered the algorithm.

Work supported by NIH Grant HD-01994 to Haskins Laboratories.

4. REFERENCES

[1] Atal, B. S., Chang, J. J., Mathews, & Tukey, J. W. (1978), "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique", *J. Acoust. Soc. Am, 63,* 1535-1555.
[2] Fant, G. (1960), *"Acoustic Theory of Speech Production"*, The Hague: Mouton & Co.
[3] Flanagan, J. L., Ishizaka, K., & Shipley, (1980), "Signal models for low bit-rate coding of speech", *J. Acoust. Soc. Am., 68,* 780-791.
[4] Kuc, R., Tuteur F., & Vaisnys, (1985), "Determining vocal tract shape by applying dynamic constraints," In *Proceedings of ICASSP* (1101-1104). Tampa Florida
[5] Levinson, S. E. & Schmidt, C. E. (1983), "Adaptive computation of articulatory parameters from the speech signal", *J. Acoust. Soc. Am., 74,* 1145-1154.
[6] Mermelstein, P. (1967), "Determination of the vocal-tract shape from measured formant frequencies", *J. Acoust. Soc. Am., 41,* 1238-1294.
[7] Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986), *"Numerical Recipes"*, Cambridge: Cambridge University Press.
[8] Shirai, K. & Kobayashi, T. (1986), Estimating articulatory motion from speech waves, *Speech Communication, 5,* 159-170.
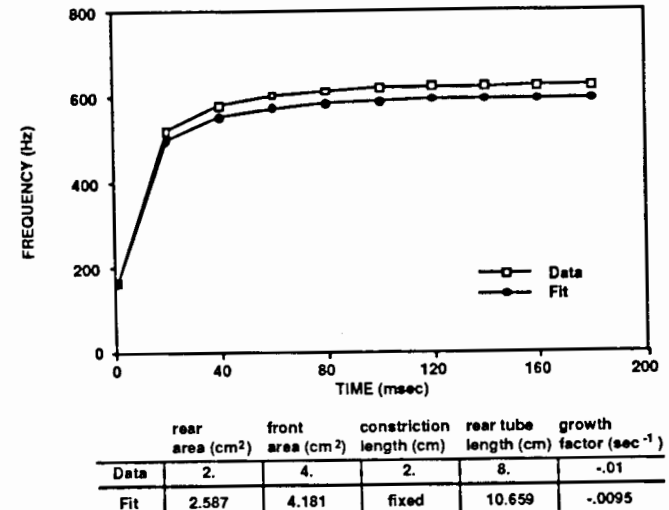
	rear area (cm²)	front area (cm²)	constriction length (cm)	rear tube length (cm)	growth factor (sec⁻¹)
Data	2.	4.	2.	8.	-.01
Fit	2.587	4.181	fixed	10.659	-.0095

Figure 4: One resonance frequency trajectory, implicit function & frequency difference minimization

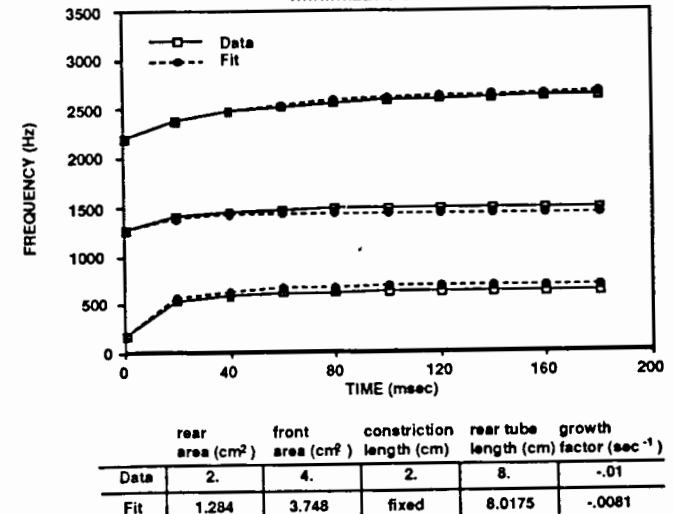|	rear area (cm²)	front area (cm²)	constriction length (cm)	rear tube length (cm)	growth factor (sec⁻¹)
Data	2.	4.	2.	8.	-.01
Fit	1.284	3.748	fixed	8.0175	-.0081

Figure 5: Three resonance frequency trajectories, implicit function & frequency difference minimization