ACOUSTIC MODELING OF TAMIL RETROFLEX LIQUIDS

Shrikanth Narayanan[1] and Abigail Kaun[2]

[1]*AT&T Labs–Research, Florham Park, NJ 07932, USA*
[2]*Linguistics Department, Yale University, New Haven, CT 06520, USA*
shri@research.att.com, abigail.kaun@yale.edu

ABSTRACT

Tamil dialects may contain as many as five contrastive liquids. This paper focuses on two of the more unusual of these liquids. The study uses MRI-derived vocal tract data in conjunction with an articulatory synthesizer to investigate articulatory-acoustic mappings in these sounds. The effects of varying constriction area, constriction location and side-cavity lengths are investigated. Simulation results show good correspondence between estimated and actual formant values.

1. INTRODUCTION

Tamil, a Dravidian language of South India, has five distinctive liquids in some of its dialects. Two of these are rhotics, including a dental (or pre-alveolar), /r/, and an alveolar, /ɾ/. A third rhotic, /ɻ/, is a palatal retroflex approximant. The other two are laterals: a dental, /l/, and a palatal retroflex, /ɭ/. A detailed characterization of the articulatory geometry and kinematics of these sounds was provided in [1]. Multiple measurement techniques – magnetic resonance imaging (MRI), magnetometry, palatography and acoustic recordings – were used to obtain articulatory and acoustic data from one native speaker of the Brahmin dialect.

MRI provides fairly accurate data regarding static vocal tract configurations. Such data have enabled better acoustic modeling of sounds that are characterized by complex vocal tract shapes such as the liquid consonants of American English (e.g., [2]). This paper focuses on the modeling of the Tamil lateral approximant /ɭ/ and the rhotic approximant /ɻ/. Both these sounds are produced as palatal retroflexes and show certain similarities in their acoustic spectra. Although /ɭ/ and /ɻ/ have similar midsagittal vocal tract profiles and oral constriction locations, only /ɭ/ has lateral diversion of flow contributing to zeros in its acoustic spectrum.

2. METHODOLOGY

Vocal tract information used for acoustic modeling was obtained using MRI from one native male speaker of the Brahmin dialect of Tamil (SN). Acoustic recordings were obtained from this subject and one other male speaker of the same dialect (SP).

2.1. MRI Data: Information about the 'static' vocal-tract shapes came from MRI scans (GE 1.5 T scanner) at contiguous 3 mm intervals in the sagittal and coronal anatomical planes, which allowed 3D views of the vocal tract to be constructed in a computer representation. Measurements of vocal tract length, area functions, and cavity volumes were also obtained. The subject, in a supine position in the scanner, produced each consonant preceded by '*pa*' (i.e., */paC/*) and continued sustaining the final consonant for about 13 seconds enabling 4 contiguous image slices to be recorded (3.2 seconds/slice). The above procedure was repeated until the entire vocal tract region was imaged. Details of image acquisition and analysis are given in [3]. Measurements of vocal tract dimensions and cavity volumes were obtained both from raw image scans and computer reconstructions of the 3D vocal tract. Area functions were obtained by resampling the 3D vocal tract at 0.43 cm contiguous intervals along, and in a plane perpendicular to, the vocal tract midline specified in a midsagittal reference image. The cross-sectional areas were computed by a pixel counting method. Vocal tract area functions (including side channels in laterals) measured from MRI data for subject SN appear in [1].

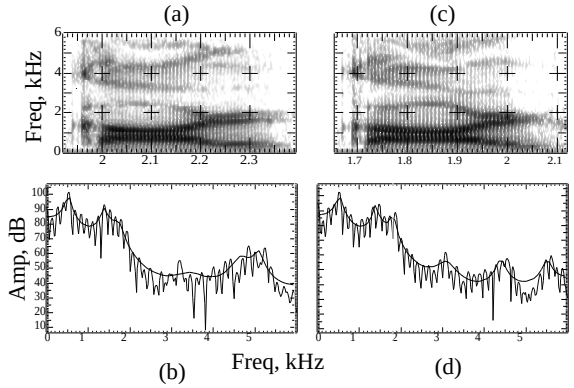

Figure 1: Spectrograms for (a) kaɭ (c) kaɻ. DFT and LPC spectra at the onset of the liquid for (b) /ɭ/ (d) /ɻ/.

2.2. Acoustic Recordings: Acoustic recordings of the liquids /ɭ/ and /ɻ/ were obtained in the following contexts: */kaCam, kiCi, kaCi, kiCam, kaC/* where *C* was {ɭ, ɻ}. Of the ten words, one was a nonsense word (kiɭam). Five repetitions of each word, embedded in the carrier phrase "*Andha vakyam* ——— *perusu*" (The phrase ——— is big), were recorded in a pseudo-random order. The subjects were two male native speakers of the Brahmin dialect (SN, SP).

Spectrograms for the words /kaɭ/ and /kaɻ/ spoken by subject SN are given in Fig. 1(a-b). Spectral slices (DFT and LPC) taken approximately at the consonant onset are shown alongside in Fig. 1(c-d). The results of the acoustic analysis are given in Table 1. The formant frequency values reported are averaged over five repetitions for each phonetic

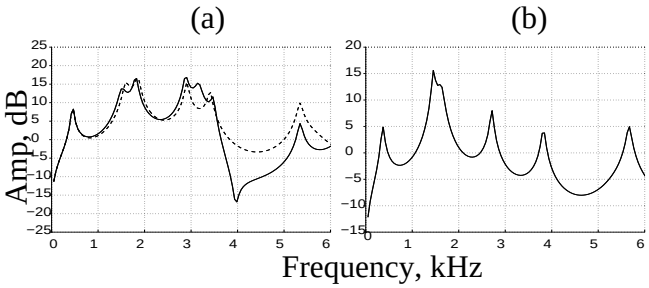

Figure 2: Transfer functions from MRI-derived area functions. (a) /ɭ/, with side cavity (solid), without (dashed). (b) /ɻ/.

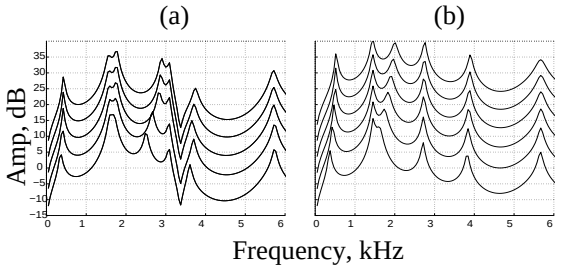

Figure 3: Effect of varying area of minimum constriction (from 0.2 - 1.2 cm^2 in steps of 0.2 cm^2) on the transfer functions. (a) /ɭ/. (b) /ɻ/. Transfer functions are plotted with a 5 dB relative shift for clarity.

Context	Formants F1	F2	F3	Zero Z
SN				
kaɭam	449	1245	2145	3106
kaɭi	320	1650	2128	3500
kiɭi	290	1609	2411	3004
kiɭam	474	1580	1948	3340
kaɭ	420	1575	1888	2750
range	290-550	1240-1740	1820-2460	2600-4000
SP				
kaɭam	573	1519	2448	3186
kaɭi	357	1651	2545	3135
kiɭi	341	1672	2606	2160
kiɭam	482	1637	2560	3320
kaɭ	460	1485	2538	3291
range	310-580	1440-1730	2325-2700	1950-3600
	F1	F2	F3	F4
SN				
kaɻam	520	1593	1913	3000
kaɻi	420	1594	1910	3060
kiɻi	300	1850	2120	2980
kiɻam	435	1649	2007	2890
kaɻ	515	1445	1845	3130
range	300-500	1440-1850	1860-2200	2845-3200
SP				
kaɻam	580	1633	2095	2953
kaɻi	447	1957	2251	3030
kiɻi	368	2167	2445	3049
kiɻam	531	1725	2200	3085
kaɻ	462	1716	2035	2808
range	350-580	1607-2200	1975-2480	2700-3290

Table 1: Formant frequency values in Hz for /ɭ/ and /ɻ/ produced by subjects SN and SP.

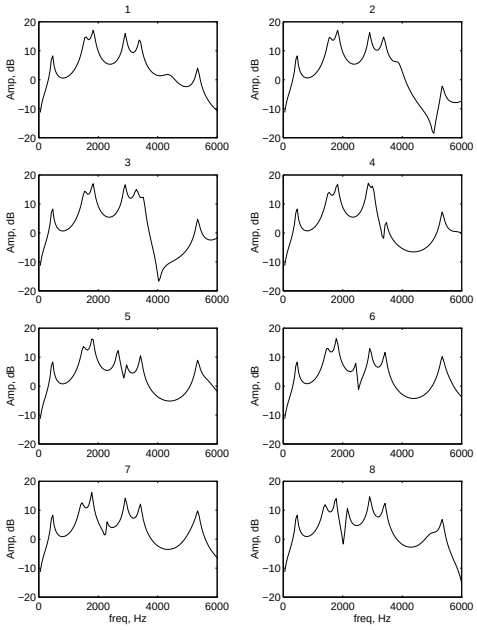

Figure 4: Effect of varying the side-cavity length for /ɭ/ from 1.72 cm (panel 1) through 4.73 cm (panel 8) in increments of 0.43 cm.

context. The range of formant frequency values across contexts is also provided for each liquid and for each subject.

In [1], the acoustic analysis was restricted to (artificially) sustained sounds from one subject (SN). For /ɭ/, F1: 400-500 Hz, F2: 1460 Hz, F3: 1800 Hz, F4: 2500 Hz, F5: 3600 Hz and a zero around 3300 Hz. For /ɻ/, F1: 400-450 Hz, F2: 1500 Hz, F3: 1850 Hz, F4: 3000 Hz. No prominent zeros were noticed below 5 kHz in the spectrum. While the results summarized in Table 1 are in general consistent with those obtained for sustained sounds, they show considerable variability across different phonetic contexts. The following observations can be made regarding the coarticulatory effects: (1) F1 frequency showed the greatest relative variability with respect to context and tended to be lower when the liquid was followed by the high vowel /i/. (2) F2 frequency tended to be lower in a low vowel context (compare results for /a_a/ and /i_i/, for example). (3) No systematic effect of context was evident in F3 frequency values. (4) For /ɭ/, the frequency of the lowest zero exhibited a wide range of values.

F1 values across the two sounds were similar. Mean F2 values of /ɻ/ across various contexts were higher than those for /ɭ/ while the opposite was true for F3 values. (F2-F3) separation was larger for /ɭ/ when compared to /ɻ/.

Across the two subjects, the range F1 of frequency values were similar. For /ɭ/, F2 frequency values were similar for both subjects while F3 values were somewhat higher in SP than SN. For /ɻ/, Subject SP has somewhat higher values for F2 and F3 frequencies compared to SN. In general, (F2-F3) separation for /ɻ/ was very similar for both subjects ((F2-F3) of formant frequency values averaged across contexts was 333 Hz for SN and 366 Hz for SP). For /ɭ/, (F2-F3) separation was subject-dependent and was greater for SP compared to SN ((F2-F3) using formant frequency values averaged across contexts was 573 Hz for SN and 947 Hz for SP).

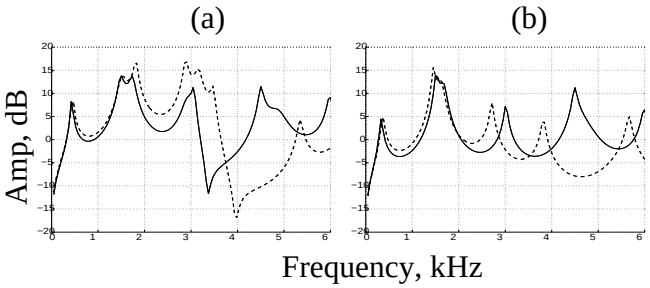

Figure 5: Comparing transfer functions generated from actual area functions (dashed) and their two-tube model approximations (solid) (a) /ɬ/. (b) /ɻ/.

3. ACOUSTIC MODELING EXPERIMENTS

In [1], basic articulatory-acoustic relations in /ɬ/ and /ɻ/ were investigated using concatenated, decoupled tube model approximations of the vocal tract area functions derived from MRI data (for subject SN). For /ɬ/, results yielded the following values: F1=540 Hz, F2=1480 Hz, F3=1822 Hz, F4=2460 Hz and Z=3340 Hz. For /ɻ/, F1=433 Hz, F2=1485 Hz, F3=1822 Hz, and F4=2970 Hz. For both /ɬ/ and /ɻ/, F1 was attributed to the Helmholtz resonance of the back cavity and constriction, F3, to the front cavity and F2 and F4, to the back cavity resonances. The zero in /ɬ/ was attributed to the side channel.

In this paper, a more realistic acoustic modeling of /ɬ/ and /ɻ/ is attempted using an articulatory synthesizer [4]. The MRI-derived area functions for subject SN will be used as the basis for the simulations.

3.1. Articulatory Synthesizer: A hybrid time-frequency domain articulatory synthesizer [4] was used to derive the vocal tract transfer functions. The wave propagation in the tract is assumed to be planar and linear and the tract is modeled in the frequency domain in terms of 2x2 chain matrices in a wave digital filter representation. The glottis is modeled in the time-domain by a nonlinear oscillator model. The implementation used for simulations reported in this paper included the ability to specify side channels and sublingual channels.

3.2. Simulation Results: Figure 2 shows the transfer functions for /ɬ/ and /ɻ/ obtained from the articulatory synthesizer. Results for /ɬ/ obtained without the inclusion of a side cavity (Fig. 2(a), dashed lines) yielded formant frequencies of F1=468 Hz, F2=1593 Hz, F3=1828 Hz, F4=2906 Hz, and F5=3422 Hz. Inclusion of a side cavity (areas listed in Appendix B, [1]) yielded a zero around 3950 Hz and an additional pole around 3140 Hz. Further, a small upward shift in F2 frequency from 1593 Hz to 1650 Hz was observed. These estimated pole frequency values are well within the range of values obtained for subject SN (Table 1). The estimated value of zero frequency is on the higher side of the range of values observed in the natural speech data of SN.

Results for /ɻ/ (Fig. 2(b)) yielded F1=380 Hz, F2=1455 Hz, F3=1595Hz, F4=2719 Hz, and F5=3845 Hz. The estimated values for F1 and F2 are within the range of values observed in the naturally spoken words of SN. However, the F3 and F4 frequency values are considerably lower than those observed in the natural speech of SN.

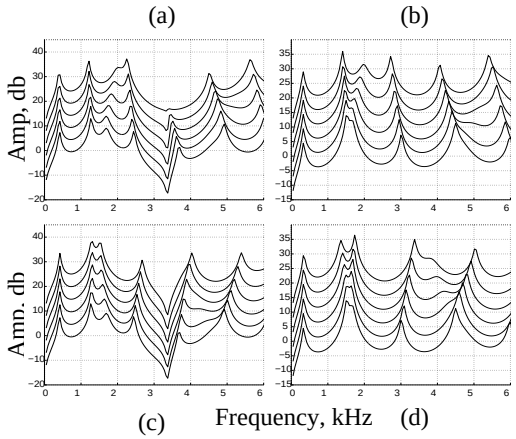

Figure 6: Effects of varying back cavity length on transfer functions from two-tube models. Increase in length from 11.61-12.86 cm in increments of 0.25 cm (a) /ɬ/ (b) /ɻ/. Decrease in length from 11.61-10.36 cm in steps of 0.25 cm (c) /ɬ/ (d) /ɻ/. Transfer functions are plotted with a 5 dB relative shift for clarity.

Note that the area functions obtained using MRI represent a sample for a single, and somewhat artificial production condition. To better understand the acoustic characteristics of these sounds, a series of simulations was performed wherein effects of varying certain aspects of the vocal tract dimensions were explored.

3.2.1. Effect of varying A_c:
The effect of varying the area of minimum constriction, A_c, is graphically shown in Fig. 3. For /ɬ/, this meant that the combined side channel opening in the region of the medial tongue closure was allowed to take a range of values between 0.2 cm^2 (bottom most curve in Fig. 3(a)) and 1.2 cm^2 (top most curve in Fig. 3(a)) in steps of 0.2 cm^2.
For /ɬ/, increasing A_c resulted in small increases in the frequencies of F1 (380→420 Hz), F3 (1700→1800 Hz) and F6 (3620→3765 Hz) and a relatively large increase in the frequency of F4 (2485→2900 Hz). There was also a small decrease in F2 frequency (1688→1545 Hz) and no noticeable changes in the frequency of the zero.
For /ɻ/, increasing A_c resulted in a modest increase in the frequency of F1 (380→515 Hz) and no changes in the frequencies of F4, F5 and F6 (Fig. 3(b)). However, increasing A_c induced a significant increase in F3 frequency (1595→2000 Hz).

3.2.2. Effect of varying side-cavity length in /ɬ/:
The effect of varying the side cavity length in /ɬ/ in the range of 1.72 cm through 4.73 cm in steps of 0.43 cm (nominal length measured using MRI was 2.58 cm) is shown in Fig. 4. As can be seen from these plots in Fig. 4, changes in the length of the side cavity has a drastic effect on the location of the zero and the overall spectral shape, particularly in the region above F3 (typically, >2 kHz). For a side-cavity length of about 1.7 cm and below, the lowest zero frequency is well above the range of our interest (>6 kHz). For lengths of 2.15 and 2.58 cm a prominent zero is seen in the region of 4-6 kHz. For side cavity lengths greater than 2.58 cm, there is a clear interaction between the pole-zero pair arising from the side-cavity and the poles due to

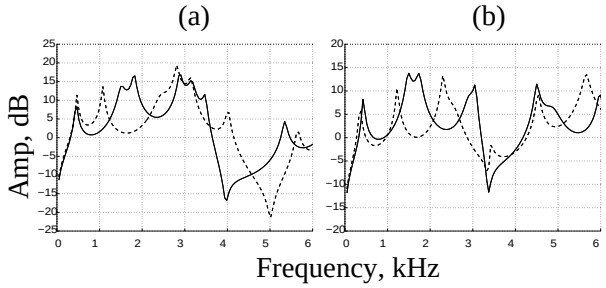

Figure 7: Comparison of transfer functions for /l/ (dashed) and /ɭ/ (solid) (a) for MRI-derived area functions (b) for two-tube approximations.

the main oral tract (frequency region of 2-4 kHz, panels 4-8 of Fig. 4).

3.2.3. Area functions vs. tube approximations

The effect of approximating the area functions (derived here from MRI data) by a simple two-tube model representing cavities in front of, and behind, the oral constriction is illustrated in Fig.5. In general, for both /ɭ/ and /ɻ/ there was minimal effect in the range below F3 frequency (< 2 kHz). However, in the range above 2 kHz, for /ɻ/ there was a general upward shift in the frequency of the higher formants F4 and F5. In the case of /ɭ/, the side cavity was also approximated by a tube of constant cross-sectional area. The most prominent change was the downward shift in the frequency of the zero by almost 600 Hz, and the concurrent effects on the overall spectrum in the region above 2 kHz.

3.2.4. Effect of varying relative cavity lengths

The effect of varying the relative lengths of the front and back cavities (equivalently, varying the constriction location) was studied using the two-tube model approximations. Moving the constriction location further forward in the mouth (i.e., increasing back cavity length and decreasing front cavity length) resulted in the following effect on the formant frequencies: (1) For /ɭ/ (Fig.6(a)), F1 and F2 did not change, F3 increased (1688→2000 Hz), F4 decreased (2438→2250 Hz), F5 decreased (3700→3350 Hz), and F6 decreased (4940→4500 Hz). The frequency of the zero was unchanged; however, the effect of the decreasing F4 approaching the value of the zero frequency tended to cancel the zero's effect. (2) For /ɻ/ (Fig.6(b)), F1 and F2 did not change, F3 increased considerably (1640-2000 Hz), F4 decreased slightly (3000→2850 Hz) and F5 decreased considerably (4500→4078 Hz).

Moving the constriction location further backward in the mouth, away from the lips, resulted in the following effects on the formant frequencies: (1) For /ɭ/, (Fig.6(c)), there were no changes in F1 and F2, a small decrease in F3 (1687→1500 Hz), significant increases in F4 (2438→2672 Hz) and F5 (3700→4030 Hz). There were no changes either in the frequency or the amplitude of the zero for the range of cavity length changes investigated here. (2) For /ɻ/, there was a small decrease in F2 frequency (1500→1360 Hz), a small increase in F3 frequency (1640→1735 Hz), and a considerable increase in the frequencies of F4 (3000→3375 Hz) and F5 (4500→5000 Hz).

3.2.5. Comparing /ɭ/ with /l/

The transfer functions for /l/ and /ɭ/ are shown in Fig.7. One major difference is in the (F2-F3) separation which is larger for /l/ compared to /ɭ/. The acoustic data for subject SN also showed a similar tendency: the (F2-F3) separation for the mean values of F2 and F3 across various contexts was 573 Hz for /ɭ/ versus 1360 Hz for /l/ (947 Hz vs. 1102 Hz for SP). The other major difference was in the frequency of the zero for the transfer functions obtained from the actual area functions: for similar side-cavity lengths, the zero frequency tends to be higher in /l/ than in /ɭ/.

4. DISCUSSION

The simulations confirm the vocal tract cavity affiliations of the various formants of /ɭ/ and /ɻ/ suggested in [1]: F3 is associated with the front cavity, F4 and F5 with the back cavity and the zero in /ɭ/ to the side cavity. Exceptions to these general articulatory associations arise when relative cavity dimensions exceed some nominal range of values: e.g., (1) There is cross-over in F2 and F3 affiliations when the relative increase in the back cavity length with respect to the front cavity length exceeds a certain value (2) Increases in side-cavity lengths in /ɭ/ beyond a certain value result in pole-zeros that effect considerable changes in the spectrum due to pole-zero canceling. Relatively high F3 frequencies in /ɭ/ and relatively high F2 frequencies in /ɻ/ seen in the acoustic data (implying a greater (F2-F3) separation in /ɭ/ compared to /ɻ/) suggest that the constriction locations may be further forward in the mouth for /ɭ/ compared to /ɻ/; larger values for A_c may also contribute to higher F3 frequencies in /ɭ/. Comparing the acoustic data of SN and SP's /ɭ/, it is likely that SP produces more anterior oral constriction locations than SN (F2-F3 separation of 573 Hz vs. 947 Hz). In this regard, it is also interesting to note the (F2-F3) separation in /l/, which is in general greater than for /ɭ/. Results for subject SP however show a smaller separation (and a tendency toward acoustic overlap). In fact, the data of /l/ and /ɭ/ suggest a tendency toward merger in this subject (informal listening to the speech tokens support this observation). This adds yet a new dimension to the discussion provided in [1] about the merger of /ɭ/ and /ɻ/ observed in certain speakers of Tamil.

The larger values for (F2-F3) separation seen in the acoustic data of SN's /ɻ/ compared to those predicted by the model may arise due to larger A_c values and/or a more anterior oral constriction location in SN's fluent speech production. The wide range of values for the frequency of the zero suggests that side cavity length is variable both across repetitions of the same token and across different phonetic contexts.

5. REFERENCES

[1] S. Narayanan, D. Byrd, and A. Kaun, "Geometry, Kinematics, and Acoustics of Tamil liquid consonants," *J. Acoust. Soc. Am., Submitted.*

[2] C. Y. Espy-Wilson, S. Narayanan, A. Alwan, and S. E. Boyce, "Acoustic modelling of American English r," in *Proc. EuroSpeech*, vol. 1, (Rhodes, Greece), pp. 393–396, Sept. 1997.

[3] S. Narayanan, A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 98, pp. 1325–1347, Sept. 1995.

[4] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. ASSP*, vol. ASSP-35, pp. 955–967, July 1987.