Acoustic Strategies for Production of American English "retroflex" /r/

Zhaoyan Zhang[†], Suzanne Boyce[‡], Carol Espy-Wilson[†], and Mark Tiede[ϒ]

[†]University of Maryland, USA

[‡]University of Cincinnati, USA

[ϒ]Haskins Laboratories, USA

E-mail: zhaoyan@glue.umd.edu, boycese@email.uc.edu, espy@glue.umd.edu, tiede@haskins.yale.edu

ABSTRACT

The variety of articulatory configurations used for American English rhotic liquids suggests that speakers used different articulatory strategies to achieve similar acoustic goals. In this paper, we focus on a variant of this liquid produced with strong retroflexion of the tongue tip and no sublingual space. Previous models of "bunched" articulatory configurations and less extreme "retroflex" versions of /r/ have ascribed the F3 resonance to the front cavity. In this paper, we propose that speakers who use the extreme retroflex version of /r/ have switched the cavity affiliation of F3 to the back cavity.

1 INTRODUCTION

Tongue configurations for the American English rhotic liquid are notoriously variable. Traditional phonetics identifies two types of tongue configuration: “bunched” and “retroflex”. Other authors have classified tongue shapes for /r/ into six types [1-2], four types [3]and three types [4]. This variability is illustrated in Fig. 1, which shows a set of midsagittal Magnetic Resonance (MR) vocal tract images, from 12 native speakers of American English producing /r/. These represent a continuum from the classic retroflex shape, with a raised tongue tip and a lowered tongue dorsum, through the type characterized as having a raised tongue dorsum and a raised tongue tip [4] to the classic “bunched” shape with a raised dorsum and a lowered tongue tip. We use the symbol /r/ for all.

The variability of /r/ articulation is particularly interesting in that all tongue configurations result in a similar acoustic profile for Formants 1, 2 and 3. In our recent work, we have been looking to explain the variability of /r/ in terms of alternate acoustic strategies. Our approach is to model the vocal tract cavities and constrictions that produce appropriate acoustic resonance profiles. Although all perceptually acceptable instances of /r/ show three constrictions, the size of these constrictions is variable. Similarly, although all instances of /r/ show three cavities, these cavities may differ in length and volume. As cavity/constriction size changes, the cavity affiliation of different formants may change as well.

Effective modeling of this type requires detailed data on the shape of the vocal tract during /r/ articulation. In particular, because cavities are modeled as quarter-wave, half-wave, or Helmholtz resonators according to the size of constrictions at endpoints, it is important to have precise data on constriction size. Recent advances in MR imaging have made it possible to acquire volumetric data on vocal tract constriction at accuracy levels around 1 mm^3. Such data, in the form of area functions, can be used to calculate the likely acoustic output of the vocal tract. This calculation can be compared with the acoustic profile of the subjects’ recorded speech. Cavity affiliations—that is, which formant arises from which cavity—can be postulated from simple-tube models based on vocal tract dimensions.

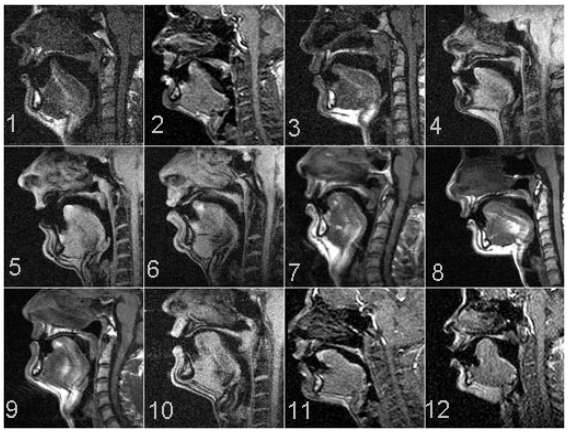

Fig. 1. American English speakers producing /r/. From Tiede (unpublished).

In previous work, we considered tongue configurations for two subjects whose vocal tract configurations during sustained /r/ showed bunching of the tongue dorsum and raising of the tongue tip. These data were derived from a study by Alwan et al. [5]. Figure 2 shows the midsagittal tongue profiles for these subjects.

The affiliation of F3 with the front cavity is universally assumed by previous models of /r/. For instance, Stevens [7] models F3 in retroflex (tip up, dorsum down) /r/ as arising from the front cavity with little lip constriction. Alwan [5] explicitly assign F3 to the front cavity, while Guenther et al. [8] note a number of front-cavity specific

ISBN 1-876346-48-5 © 2003 UAB

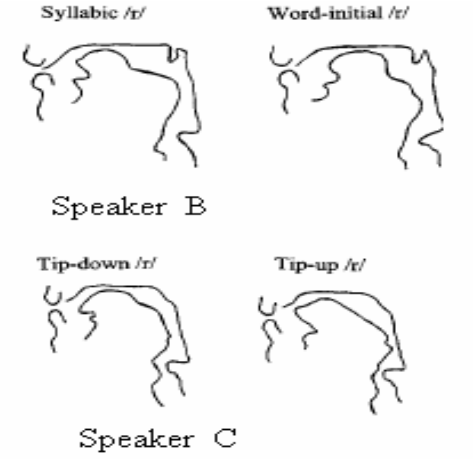

Fig. 2. Midsagittal tracings of vocal tract configurations used as model input in [6]. Speaker B was instructed differently than Speaker C [5].

trading relations for F3. In this model, F3 comes from the front cavity, while F2 and F1 come from the mid and back cavities respectively.

F3 for the speakers of Fig. 2 covered a range of 1400-2100 Hz [6]. Front cavity resonances of this magnitude ordinarily occur in back vowels with palatovelar constrictions, and are normally the second resonance in the spectrum rather than the third. In this model, the front cavity is a Helmholtz resonance due to the difference in area between the lip constriction and the front cavity. Because of the anterior palato-alveolar constriction, the addition of the sublingual space to the front cavity volume is instrumental in achieving an F3 in the appropriate range.

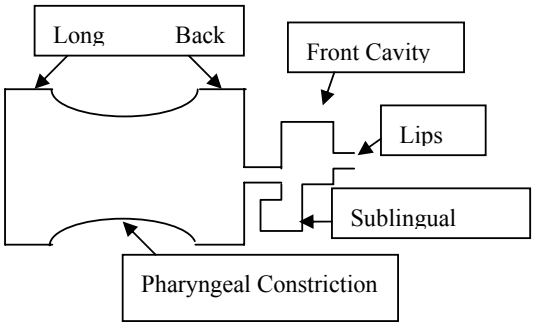

Fig. 3. Schematic illustration of simple-tube model from Subject C in Fig. 2 producing a tip-up bunched /r/. All dimensions are in centimeters. Glottis at left, lips at right. From Espy-Wilson et al. [6].

The model of Espy-Wilson et al. [6] was quite successful in predicting formants that matched subjects' actual speech for the tongue configurations shown in Fig. 2. (The model was only moderately successful for the tip-down configuration of Speaker C, for reasons that were not clear.) For each of these configurations, the front cavity was the only possible source for F3. Further, without the addition of the sublingual space the front cavity volume was not sufficient to predict an F3 in the appropriate range. For the tip-down /r/, with negligible sublingual space, the increased volume came from a more posterior constriction along the palate. This may reflect a trading relationship between constrictions.

In this paper, we turn our attention to modeling the retroflex /r/ of Speaker 1 in Fig. 1. This /r/ differs in some particulars from the classic retroflex /r/ as modeled by Stevens [7] below). First, Stevens assumes a rather large sublingual space. Such a space is not visible as a separate cavity for Speaker 1. Rather, the underside of Speaker 1's tongue remains perpendicular to the floor of his mouth for most of its extent. Further, Stevens assumes a constriction at the alveolar ridge. Clearly, the constriction for Speaker 1 between the tongue and the palate is somewhat posterior to the alveolar ridge. These different cavity shapes and sizes suggested that a different model was needed for Speaker 1's retroflex /r/.

2 RESULTS

For the purpose of acquiring vocal tract dimensions in terms of area functions, a semi-polar grid superimposed on a midsagittal projection through the MRI volume dataset was first aligned using anatomical landmarks [10]. The centroids of the intersection of this grid with the anterior and posterior tract walls were then used to define the approximate airflow path through the tract. Finally, cross-sections orthogonal to this path projected through the volume were thresholded to obtain measures of area at each centroid offset. Measured cross sectional area functions from the MR images of Speaker 1 are shown in Fig. 4.

Acoustical data were also recorded from Speaker 1, but at a time independent of the MR imaging session. To replicate the supine posture required for MR imaging, these data were recorded with the subject in supine position. Fig. 5 below shows the power spectrum of the acoustical data. The first formants are obtained from the figure and shown in Table 1.

The measured area functions were used as input to a vocal tract modeling program modified from the VTCALCS program of [11]. These modifications are described in [12] and [13]. The resulting estimates of F1 - F5 values are also shown in the second row of Table 1. The agreement with the real values is very good.

As a first step, we substituted the dimensions measured from Speaker 1 into the simple-tube model of [6]. The vocal tract is divided into six sections. From the glottis to the lip, they are: back cavity, pharyngeal constriction, middle cavity, oral constriction, front cavity, and lip constriction. The area for each section was obtained by averaging the area function over its length. The correspondence between simple-tube dimensions and measured dimensions can be seen in Fig. 4.

The calculated formant frequencies from the simple tube model can be found in the third row of Table 1. As can be readily seen, estimates of F1-F3 from the measured and simple-tube area functions are nearly identical with actual

formant values and with each other. Given the uncertainties involved in obtaining vocal tract dimensions, and the uncertainties involved in simple-tube assumptions, agreement in the vicinity of 100 Hz is extremely good. The estimates were somewhat less successful for F4 through F6, with a maximum gap of 265 Hz for F4. This is probably due to the simplifications made in obtaining the simple-tube dimensions.

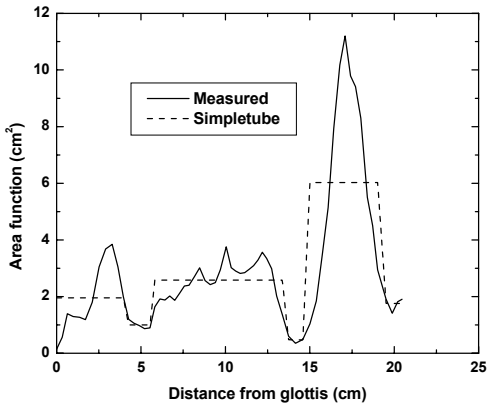

Fig. 4. **Measured area function and simple-tube area function.**

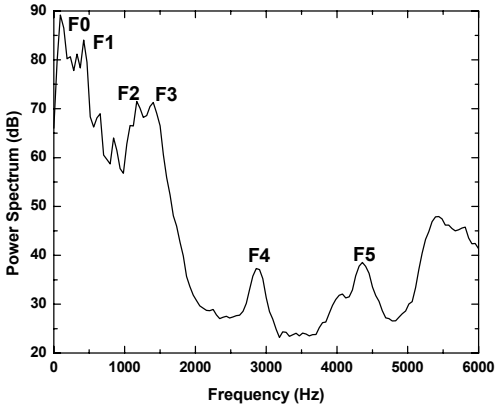

Fig. 5. Power spectrum of the /r/ produced at supine position.

The cavity affiliation of formants was investigated by estimating formant frequencies from the simple-tube model [6]. For each cavity, this endeavor requires an explicit decision regarding the appropriate type of resonator (i.e. quarter-wave, half-wave, etc.) to assume. In some cases, the dimensions made it obvious. In other cases, we modeled the cavity in several alternative ways. The most successful is discussed below.

Speaker 1 has a relatively weak constriction in the pharynx compared to some of the other speakers in Fig. 1. In [6], we made the point that if the constriction in the pharynx is narrow enough, the long back cavity must be modeled as two decoupled cavities separated by a constriction. In the case of Speaker 1, estimates based on assuming decoupled cavities were less successful in predicting formant values.

	F1	F2	F3	F4	F5
Real formant value	420	1170	1380	2860	4360
Prediction from measured area function	440	1020	1360	2840	4395
Prediction from simple-tube area function	455	1045	1420	2595	4120
Estimation from simple-tube model	454	1089	1274	2547	4160

Table 1. Comparison between the predicted and real formants.

Accordingly, the area of the vocal tract posterior to the oral constriction was modeled as a single cavity with a perturbation. F1 was estimated as the Helmholtz resonant frequency of the long back cavity plus the oral constriction. The relative difference in area between the front cavity and lip constriction makes it reasonable to model the front cavity as a Helmholtz resonator. We modeled F3 and F4 as the first and second half-wavelength resonance of the long back cavity, respectively. F2 was the Helmholtz resonant frequency of the front cavity plus the lip constriction, while F5 was the first half-wavelength resonant frequency of the front cavity, with correction for the lip opening. Values from this model are reported in the last row of Table 1.

In summary, we model cavity affiliations as follows:

F1: Helmholtz resonance of the cavities posterior to the oral constriction and the oral constriction;

F2: Helmholtz resonance of the front cavity and the lip constriction;

F3: first half-wavelength resonance of the long back cavity;

F4: second half-wavelength resonance of the long back cavity;

F5: first half-wavelength resonance of the front cavity with correction for lip opening. (Alternatively, it could be due to the first half-wavelength of the back cavity, as the back cavity and front cavity have similar length.)

3 DISCUSSION

The major difference between the retroflex /r/ of Speaker 1 and the bunched /r/ modeled in [6] is the change in cavity affiliation of F2/F3. This comes about because the posterior position of the constriction between tongue and palate, together with the shorter length of the constriction (1.26 cm compared to 3 cm in [6]), have the effect of enlarging the front cavity to produce a large Helmholtz resonator. The effect is to lower the Helmholtz frequency of the front cavity into the F2 region, leaving the frequency

ISBN 1-876346-48-5 © 2003 UAB

of the first half-wavelength resonance of the long back cavity to emerge as F3. These effects cause the change in cavity affiliation of F2/F3, as compared with that for bunched /r/. In contrast to our findings for the raised dorsum "bunched" /r/'s discussed in [6], Speaker 1's large front cavity obviates the need for an extra sublingual cavity. It seems that the simple-tube model discussed in this study can be used for both bunched and retroflex /r/ sounds.

The finding that F3 arises from the back cavity rather than the front cavity for Speaker 1's /r/ is somewhat iconoclastic, in that (as noted above) the assumption that F3 rises from the front cavity is firmly entrenched in the literature. Interestingly, Speaker 1's formant values for F2 and F3 are extremely close to each other, with F3 appearing on the power spectrum as a shoulder on the slope of F2. In the model, they are also extremely close, suggesting that cavity affiliations would switch again with a slightly different constriction placement or area. Among normal speakers of American English producing /r/, there is a substantial subset whose F2 and F3 come close enough to merge. It is tempting to speculate that other speakers whose /r/'s show overlapping F2 and F3 values, may use a tongue configuration similar to that of Speaker 1.

In his discussion of raised dorsum "bunched" /r/, Stevens [7] discusses possible roles for side channels around the tongue. The good agreement between estimated and real formant values in this study suggests the contribution of such air paths, if any exists, should be insignificant.

These findings point out the unique nature of American English /r/, in that functionally equivalent acoustic profiles can be produced with very different vocal tract configurations. Further, these different vocal tract configurations employ very different acoustic strategies to match the formant values of /r/. It is hard to think of an alternative case where cavity affiliation can change without affecting the viability of the acoustic output. We plan to explore this issue further in future work.

4 ACKNOWLEDGEMENTS

This work was supported by NIH grant 1 R01 DC05250-01 to Suzanne Boyce, Carol Espy-Wilson and Mark Tiede, and by NIH grant 5-K02 DC00149-03 to Carol Espy-Wilson. As always, we acknowledge the inspiration of Ken Stevens.

5 REFERENCES

[1] Delattre, P. and D. Freeman, "A dialect study of American English R's by X-ray motion picture", *Language*,. vol. 44: pp. 28-69, 1968.

[2] Hagiwara, R., "Acoustic Realizations of American /r/ as Produced by Women and Men". *UCLA Phonetics Laboratory Working Papers*, vol. 90**,** 1995.

[3] Westbury, J.R., Hashi, M. and Lindstrom, M.J., "Differnces among speakers in lingual articulation for American English /r/". *Speech Communication*, vol. 26: pp. 203-226, 1998.

[4] Boyce, S. and C.Y. Espy-Wilson, "Coarticulatory Stability in American English /r/". *Journal of the Acoustical Society of America*, vol. 101(6), pp. 3741-3753, 1997

.[5] Alwan, A., S. Narayanan, and K. Haker, "Toward Articulatory-acoustic Models for Liquid Approximants based on MRI and EPG data. Part II: The Rhotics". *Journal of the Acoustical Society of America*, vol**.** 101: pp. 1078-1089,1997.

[6] Espy-Wilson, C.Y., S. Boyce, M.T.T. Jackson, S. Narayanan, & A. Alwan," "Acoustic Modeling of American English /r/," *Journal of the Acoustical Society of America*, vol. 108(1): pp. 343-356, 2000.

[7] Stevens, K.N., *Acoustic Phonetics*. Cambridge: M.I.T. Press, 1998.

[8] Guenther, F.H., C.Y. Espy-Wilson, S.E. Boyce, M. L. Matthies, M. Zandipour, and J.S. Perkell., "Articulatory Tradeoffs Reduce Acoustic Variability During American English /r/ Production". *Journal of the Acoustical Society of America,* vol**. 105**: pp. 2854-2865, 1999.

[9] Perkell J.S.,.M.L. Matthies, M.A. Svirsky and M. I. Jordan, "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study". *Journal of the Acoustical Society of American,* vol 92: pp. 2948-2961, 1993.

[10] Tiede, M., H. Yehia, and E. Vatikiotis-Bateson, "A shape-based approach to vocal tract area function estimation". *Proceedings ETRW-SPM96 (Autrans*), pp. 41-44,1996.

[11] Maeda, S., "A Digital Simulation Method of the Vocal-Tract System". *Speech Communication*, vol 1,: pp. 199-229, 1982.

[12] Jackson, M., C. Y. Espy-Wilson., and S. E. Boyce, "Verifying a vocal tract model with a closed side branch". *Journal of the Acoustical Society of America*, vol 109 (6), pp. 2983-2987, 2001.

[13] Zhang, Z. and C. Y. Espy.-Wilson., "A vocal tract model with parallel side channels", manuscript in preparation, 2003.