ISCA Archive
http://www.isca-speech.org/archive

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

ACOUSTIC MODELLING OF AMERICAN ENGLISH /r/

Carol Y. Espy-Wilson[1,5], *Shrikanth Narayanan*[2], *Suzanne E. Boyce*[1,3,5] & *Abeer Alwan*[4]

[1]Dept. of Electrical and Computer Engineering, Boston University, Boston, MA 02215

[2]AT&T Labs, Murray Hill, NJ

[3]Dept. of Communication Sciences and Disorders, University of Cincinnati

[4]Dept. of Electrical Engineering, University of California at Los Angeles

[5]Research Laboratory of Electronics, Massachusetts Institute of Technology

E-mail: espy@formant.bu.edu, shri@research.att.com,boyce@formant.bu.edu, alwan@icsl.ucla.edu

ABSTRACT

The low F3 of American English /r/ (typical range 1300-1900 Hz) is accompanied articulatorily by constrictions in the pharyngeal, palatal and labial regions. Because acoustical theory predicts that formants will lower at points of maximum volume velocity in the vocal tract, and because such points occur in the pharyngeal, palatal and labial regions, many investigators have speculated that the combination of these constrictions accounts for the low F3 of /r/. In this paper, we use the Maeda vocal tract modelling software to compare theoretical predictions of constriction location to data gathered from two American English speakers via Magnetic Resonance Imaging (MRI). We conclude that additional mechanisms are required to explain the acoustics of American English /r/.

INTRODUCTION

American English /r/ is sometimes cited as an example of a many-to-one relationship between articulatory configurations and acoustic results. Speakers of rhotic dialects of American English use a multitude of different articulatory configurations for /r/ [4,14,8,13,1]. Typically, these articulatory configurations include constrictions in the pharynx, along the palatal vault, and some degree of constriction at the lips . Figure 1 (taken from [1]) shows mid-sagittal MRI tracings from some attested examples of speakers producing /r/ variants. Each of these articulatory configurations gives rise to a distinctive characteristic of /r/-- extremely low F3 values, often close to F2 [5,9,7,6]. A typical example of intervocalic /r/ is shown spectrographically in Fig. 2. As in this case, F3 may almost merge with F2 into a single formant with a wide bandwidth. In a survey of formant values reported in the literature, we found typical ranges across speakers of 250-550 Hz for F1, 900-1500 Hz for F2, and 1300-1950 Hz for F3 (2,7,6,13,9].

According to a well-known provision of the acoustic theory of the vocal tract, known as Perturbation Theory, standing wave patterns give rise to points in the vocal tract where volume velocity is at a maximum. Constriction of the vocal tract at these points will have the effect of lowering formant frequencies. Assuming a uniform tube, there are three points along the vocal tract where a constriction will cause F3 to lower. These correspond to l/5, 3l/5 and 5l/5 where l = vocal tract.

Fig. 1 Vocal tract profiles for /r/ (from [1]).

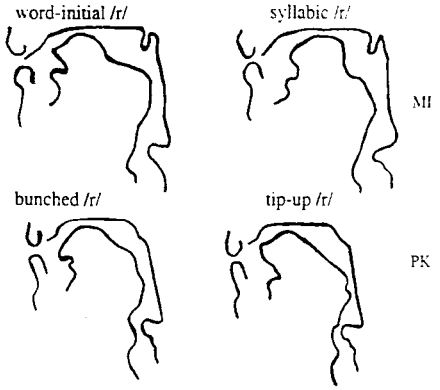
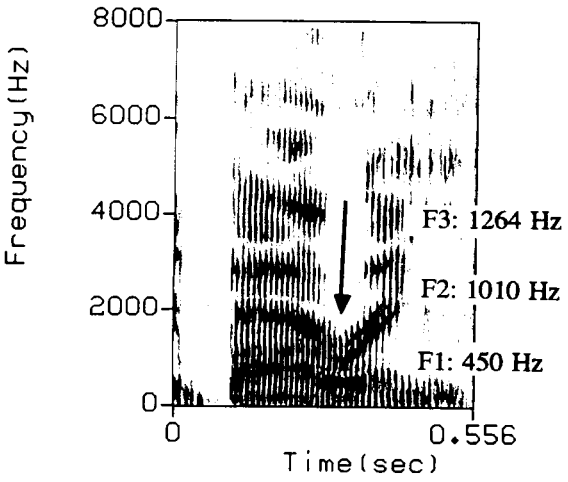


Fig. 2. Spectrogram of /barring/ (male speaker).



length. Similarly, there are three points along the vocal tract where pressure is at a maximum. These correspond to 0, 2l/5 and 4l/5. For each point, the effect of a constriction on F3 is predicted to be small--in the range of 100-200 Hz. Between points of maximum pressure and maximum volume velocity there exists a continuum of effects on F3. Perturbation Theory is most applicable

for non-obstruents whose degree of constriction permits coupling between cavities. [3,15]

Because standing wave patterns predict that points of maximum volume velocity will occur in the pharyngeal, palatal and labial regions, and because the common denominator across various types of /r/ is the presence of constrictions in these regions, it has been assumed that speakers make use of these points to lower F3 for /r/. The variety in types of /r/ seen has been attributed to idiosyncratic combinations of constrictions, each of which may individually lower F3 (10,8). This approach assumes (1) that subjects' constriction locations will coincide with the points of maximum velocity predicted by Perturbation Theory, and (2) that the net lowering achieved by each constriction can sum or otherwise combine with that achieved by others to produce the degree of F3 lowering characteristic of /r/. To date, investigators have not been explicit about the relative contribution of the different constrictions to F3 lowering, nor have they explicitly discussed the acoustical theory involved.

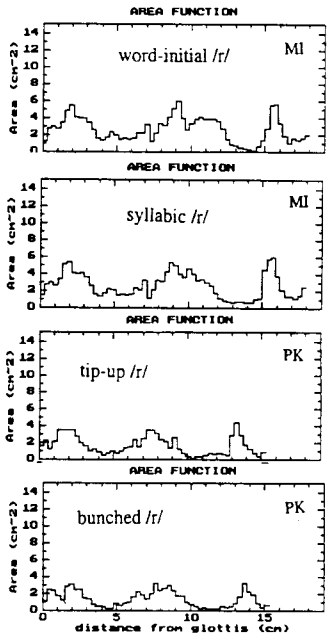
More specific acoustic models of /r/ can be found in Alwan et al. [1] and Stevens [12]. In these models, formants are derived separately from back, mid and front cavities together with constriction regions. According to Alwan et al. (1), F1 may be a Helmholz resonance formed by the palatal constriction and the cavity posterior to it; (2) F2 may be due either to helf-wavelength resonance of the cavity between the palatal and pharyngeal constrictions, or to a Helmholtz resonance formed by the pharyngeal constriction and the cavity posterior to it; and (3) the front cavity between the lips and the palatal constriction probably gives rise to F3. Because longer cavities give rise to lower formants, this model accounts for the extreme low values of F3 typical of /r/ by positing a long front cavity. Articulatorily, this may be accomplished by moving the palatal constriction back in the mouth, or (to a lesser extent) by protruding the lips. The space formed below the tongue when the front of the tongue is lifted up may also contribute to the front cavity. Alwan et al. [1] suggest that the sublingual space acts to increase the volume of the front cavity; the sublingual space may also act as a separate resonator, giving rise to a resonance/antiresonance pair [12].

In this paper, we are attempting to account for the range of F3 values seen across speakers in the literature as well as use measurements of vocal tract dimensions derived from 3-dimensional Magnetic Resonance Imaging (MRI) data (1) as input to an acoustical modelling study of the two models described above. We concentrate on 2 subjects from this study, one male and one female. The female subject (a trained phonetician) produced sustained /r/ with two different articulatory configurations. In the one case, she was instructed to produce an /r/ with tongue tip up, and in one case with tongue tip down. The male subject was instructed to produce sustained /r/ as in syllable-initial position, and as a syllable nucleus. These conditions produced 4 sets of vocal tract dimensions, corresponding to vocal tract profiles in Fig. 1. (Note that where the tongue tip was raised, supralingual and sublingual area functions were measured separately. Sublingual area functions were not included in these data.) In addition, for each subject's vocal tract we calculated standing wave points where pressure and volume velocity should be maximal. If the subjects are taking advantage of maximum volume velocity points to achieve lowered F3 during /r/, their attested constriction locations should correspond to the predictions of perturbation theory. In addition, if different combinations of constrictions have equivalent effects on F3, we would expect modelling studies based on corresponding vocal tract dimensions to produce plausibly similar acoustic profiles for /r/. Results are compared to subjects' actual formant frequencies measured from 11 /r/-containing real words produced 4 times each in a carrier phrase during a separate recording session. Formant frequencies were measured at the lowest point of F3.

Formant Values Generated from MRI Data

MRI data for each sustained /r/ were converted to area functions appropriate for the Maeda (1980) vocal tract model VTCALCS, a computer program using standard acoustical tube assumptions to predict formant frequencies from vocal tract dimensions. For Subject PK (female) this involved a model vocal tract of 15.81 cm, divided into 51 sections of .31 cm length. For Subject MI (male), this involved a model vocal tract of 18 cm, divided into 60 sections of .3 cm length. These MRI-derived values were converted to area functions in a format appropriate for input to VTCALCS. Fig. 3 shows VTCALCS area functions for each of the 4 data sets. The data are positioned such that the glottis is to the left, at 0 cm, while the lip opening is on the right. Larger areas under the curve indicate vocal tract cavities; constrictions are shown by the distance between the data curve and the abscissa.

Fig. 3. Area Functions for VTCALCS



Area functions from each data set were then used as input to predict formant values using VTCALCS. These predicted formant values are shown in Table A.

Table. A. F1-F3 values predicted by VTCALCS from MRI data.

	PK tip up	PK tip down	MI initial	MI syllabic
F1	440 Hz	440 Hz	376 Hz	400 Hz
F2	1200 Hz	1354 Hz	1194 Hz	1226 Hz
F3	2173 Hz	1952 Hz	1929 Hz	1904 Hz

Due to noise in the experimental chamber, it was not possible to record speakers' acoustic output during the MRI sessions. However, each subject produced 4 repetitions of 9 or 11 real words containing initial or syllabic /r/. Table. B shows the mean, range and standard deviation of formant frequencies produced during these real words. As the table shows, each subject's values cover a wide range of F3. Comparison of the subjects' acoustic data to predicted F3 values generated from VTCALCS shows the VTCALCS output to be acceptable (at 1952 or 2173 Hz for PK and approximately 1900 Hz for MI) but rather high compared to subjects' mean F3 and range (Range = 1479-2157 Hz for PK, 1186-1946 Hz for MI). Alternatively, there may be a need to incorporate the sublingual space into the VTCALCS model.

Table. B. Measured F1-F3 Values from Recorded Speech

Measured F1-F3	PK (N=36)	MI (N=44)
Mean F1	350 Hz	388 Hz
Mean F2	1355 Hz	1384 Hz
Mean F3	1834 Hz	1660 Hz
Range of F3	1479-2157 Hz	1186-1946 Hz
St. Dev F3	138 Hz	141 Hz

"Real" vs. "Predicted" Constriction Locations

The MRI-derived dimensions were also used as a guide to construct uniform vocal tracts appropriate to each subject. Because MRI data showed the maximum area over the vocal tract for PK to be approximately 4 cm, and that for MI to be approximately 5.5 cm, the tubes were uniformly 4 or 5.5 cm in area. F3 values for this uniform (neutral) vocal tract were 2304 Hz for subject MI and 2720 Hz for subject PK. Points of maximal volume velocity, where constriction decreases F3, and maximal pressure, where constriction increases F3, were also calculated for each subject's vocal tract. These were expressed in terms of number of cm from the glottis, where 0 = the glottis. Actual ranges across which constriction is maximal, as shown in area functions of Fig. 1, were compared with predicted constrictions according to perturbation theory. The criterion for constriction beginning and end was set at Area =1.0 cm^2 for PK and at Area =2.4 cm^2 for MI. Wide ranges indicate stretches for which constriction was similarly narrow.

Table C. Real vs. Predicted Constriction Locations

	PK tip up	PK tip down	MI initial	MI syllabic
Actual Palate	10.2-13.3 cm	10.5-13.6 cm	12.6-15.0 cm	12.0-15.0 cm
Predicted Palate	9.6 cm	9.6 cm	10.8 cm	10.8 cm
Actual Pharyx	4.3 -6.2 cm	4.3-5.9 cm	3.9-6.9 cm	3.9-6.3 cm
Predicted Pharyx	3.1 cm	3.1 cm	3.6 cm	3.6 cm

It is clear that for both speakers, real palatal constrictions are long, and appear to cover areas considerably forward of the constriction location predicted by Perturbation Theory to have the maximal lowering effect on F3. Indeed, for both subjects palatal constrictions center over areas predicted by Perturbation Theory to correspond with maximal pressure, making constrictions in these areas more likely to raise F3 than to lower it. For PK, the pharyngeal constriction ranges forward of the predicted point. For MI, the pharyngeal constriction apparently covers an area that may be conducive to F3 lowering, but is longer than necessary.. Thus, neither the pharyngeal nor the palatal constriction are located as would be predicted by Perturbation Theory. In particular, the palatal constrictions here are in areas that should affect F3 in the wrong direction. This is true regardless of the type of /r/; for instance, PK's "tip-down" /r/ and "tip-up" have slightly different constriction length but similarly forward constriction locations and similar constriction degrees. We conclude that subjects are not taking advantage of points of maximum volume velocity along the vocal tract to lower F3 in any obvious way.

Effects of Changing Constriction Location

In order to contrast the effect of actual constriction locations with the effect of constrictions at the locations predicted by Perturbation Theory, we constructed sets of vocal tract area functions with maximum areas of 4 or 5.5, as in the uniform tube described above, but with constrictions at the lips, pharynx and palate that followed closely in degree, length and gradient the subjects' actual constrictions (as shown in Fig. 1). For this work, we used the more extreme constriction types shown by MI for syllabic /r/, and by PK for "tongue-tip up" /r/. These area functions, with (1) constrictions placed as in measured MRI data, (2) centered in locations predicted by Perturbation Theory, and (3) at sections in between, were then input to VTCALCS. Table D shows formant frequency results for the endpoints of the continuum --i.e. the (Perturbation Theory) predicted and real constriction locations. It is clear that, given similar constriction degree, length and gradient, F3 has the potential to be lower in the location predicted by Perturbation Theory. In fact, at that point, F3 is close to the lowest point measured for each subject's real words. In subsequent modelling trials, we discovered that at the real palatal constriction location,

manipulations of additional factors (1) constriction degree, (2) constriction gradient, and (3) constriction length had little effect. In other words, at the real constriction location, F3 remained high relative to the subjects' ranges, and relatively intractible.

Table D. Real Constriction Location vs. that predicted by Perturbation Theory, and Removal of pharyngeal constriction

Subject MI

Vtcalcs Results	Predicted Location	Real Location	Real Location w/o Ph. Cnst.
F1	376 Hz	336 Hz	296 Hz
F2	1040 Hz	945 Hz	1506 Hz
F3	1312 Hz	1964 Hz	1984 Hz

Subject PK

Vtcalcs Results	Predicted Location	Real Location	Real Location w/o Ph. Cnst.
F1	464 Hz	416 Hz	360 Hz
F2	1249 Hz	1315 Hz	1792 Hz
F3	1702 Hz	2033 Hz	2144 Hz

An additional point to be made concerns the role of the front cavity, which is naturally longer when the palatal constriction is moved to the position predicted by Perturbation Theory. According to Alwan [1] and Stevens [12], F3 is primarily a front cavity resonance. To test this hypothesis, we removed the pharyngeal constriction entirely from the vocal tract and replaced it with a portion of the uniform tube. As the final column of Table D shows, removing the pharyngeal constriction has only minimal effects on F3 (but causes major effects in F2). Further, the fact that F3 is high at the real constriction location suggests either that the front cavity is too short, or that some other acoustic effect is in operation. The fact that only manipulations of constriction location, rather than gradient, length or degree, affect F3 to any considerable degree suggests likewise.

CONCLUSION

Overall, the following three points appear clear: (1) that the subjects' actual palatal (and pharyngeal to a lesser extent) constrictions are further forward in the vocal tract than predicted by Perturbation Theory, (2) that the F3 values predicted from real constriction locations appear to be high, (3) that F3 is associated with the front part of the vocal tract, and (4) that the full range of F3 values produced by the speakers cannot be accounted for as a simple function of front cavity length. At first glance, then, it appears that Perturbation Theory, by itself, cannot account for the acoustics of /r/. This is not by itself surprising; Perturbation Theory is more indicative of direction of change than amount of change, and is most appropriate for lesser constrictions. Note, however, that the MRI data treated here do not include the sublingual space. To account for the full range of F3 it may be necessary to (1) model the front cavity as a Helmholz resonator, and/or (2) include the sublingual space. Modelling the front cavity as a Helmholz resonator may be justified given the tapering gradient of the teeth and lips, with or without rounding. The sublingual space may act to increase the volume of the front cavity and thereby lower F3. Stevens [12] suggests a mechanism whereby the sublingual space acts as a branch cavity, setting up an additional resonance/antiresonance pair. In future work we plan to explore each of these options as possible acoustic sources of formant values in American English /r/.

REFERENCES

[1] A. Alwan, S. Narayanan & K. Haker, "Toward Articulatory-acoustic Models for Liquid Approximants based on MRI and EPG data. Part II: The Rhotics", *J. Acoust. Soc. Am.*, Vol 101, pp. 1078-1089, 1997.

[2] S. Boyce & C. Y. Espy-Wilson, "Coarticulatory Stability in American English /r/, *J. Acoust. Soc. Am.*, Vol. 102, in press, 1997.

[3] T. Chiba & M. Kajiyama. "The vowel: Its nature and structure", Kaiseikan, Tokyo, 1941.

[4] P. Delattre, & D. Freeman, ``A Dialect Study of American R's by X-ray Motion Picture," *Language*, Vol. 44, pp. 29-68, 1968.

[5] C. Espy-Wilson, ``Acoustic Measures for Linguistic Features Distinguishing the Semivowels in American English," *J. Acoust. Soc. Am.*, Vol. 92, pp. 736-757, 1992.

[6] R. Hagiwara, ``Acoustic Realizations of American /r/ as produced by Women and Men," *UCLA Phonetics Laboratory Working Papers*, Vol 90, 1995.

[7] I. Lehiste, ``Acoustical Characteristics of Selected English Consonants," *University of Michigan Communication Sciences Laboratory Report* #9, 1962.

[8] M. Lindau & P. Ladefoged, ``Variability of feature specification," in *Invariance and Variability in Speech Processes* ,edited by J. Perkell & D. Klatt, (Laurence Erlbaum, Hillsdale NJ), pp. 464-479, 1986.

[9] F. Nolan, *The phonetic bases of speaker recognition*, Cambridge University Press, Cambridge, England, 1983.

[10] J. Ohala, "Around Flat", in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin, (Academic Press, Orlando,FL), 1985.

[11] J. Olive, A. Greenwood & J. Coleman, *Acoustics of American English speech*. Springer-Verlag, New York, NY, 1993

[12] K. N. Stevens. "Acoustic Phonetics", M.I.T. Press, Cambridge, MA, In press, 1997.

[13] J. Westbury, M. Hashi & M. Lindstrom, "Differences among speakers in articulation of American English /r/: An x-ray microbeam study. In Proceedings of the XIIIth International Conference on Phonetic Sciences, August 1995.

[14] P. Zawadaski & D. Kuehn, ``A cineradiographic study of static and dynamic aspects of American English /r/," *Phonetica* Vol. 37, pp. 253-266, 1980.

[15] V. Zue. *Speech Spectrogram Reading*. Summer Course, MIT, 1985.