A SIMPLE TUBE MODEL FOR AMERICAN ENGLISH /r/

Carol Espy-Wilson* and Suzanne Boyce**

*Electrical and Computer Engineering, Boston University, Boston MA, USA
**Communication Sciences and Disorders, Univ. of Cincinnati, Cincinnati OH, USA

ABSTRACT

For American English /r/, a large number of articulatory configurations produce essentially equivalent acoustic profiles of F1, F2 and F3. In previous work, recently acquired physiological data from Magnetic Resonance Imaging studies were used to develop a simple tube model for /r/ that gives reasonable estimates for the first four formants. All the /r/ configurations showed bunching of the tongue dorsum. Using articulatory data from several sources, we show in this paper that the simple tube model also gives reasonable estimates of F4 and F5 which can vary substantially depending upon the tongue shape. Our acoustic data show that the traditional “retroflex” /r/ (with the tongue tip raised and the tongue dorsum lowered) has an acoustic profile where F3 and F5 track each other. In this case, there is also some lowering in F4. However, when the tongue dorsum is raised (regardless of whether the tongue tip is raised or lowered), there is no such lowering of F4 or F5. With reasonable estimates of vocal tract dimensions, we show that the model can explain both types of formant patterns.

1. INTRODUCTION

American English /r/ is often cited as an example of a many-to-one articulatory-acoustic relationship. Speakers of rhotic dialects of American English use a multitude of different articulatory configurations for /r/ (c.f. [1]). Overall, articulatory configurations for /r/ involve three constrictions: in the pharynx, along the palatal vault, and at the lips. Fig. 1 (adapted from [2]) shows midsaggital MRI tracings from some attested examples of speakers producing /r/ variants.

![Fig. 1. Acoustic profiles for /r/ taken from [2].](image)

The configurations differ most by what happens in the palatal region, i.e. by whether the effective constriction occurs (1) at the alveolar ridge and is made solely by the tongue tip, (2) in the palatovelar region and is made solely by the tongue dorsum with a lowered tongue tip, or (3) in both alveolar and palatovelar regions, and is made by the simultaneous raising of the tongue tip and tongue dorsum. Researchers so far have failed to link patterns of acoustic variability in formant values with the different articulatory configurations.

Traditionally, these configurations have been divided into contrasting categories of “retroflex” (in which the tongue tip is raised and the tongue dorsum is lowered) vs. “bunched” (in which the tongue dorsum is raised and the tongue tip lowered) [1,3]. However, as a number of researchers have pointed out, these two categories are only the extremes in a continuum that includes many incremental variants [1,4,5,6]. Further, the variant in which both tongue tip and tongue dorsum are raised does not easily fit into the traditional dichotomy of “retroflexed” vs. “bunched”. In this paper, therefore, the major types of configurations will be categorized as (1) tip-up retroflex /r/, (2) tip-up bunched /r/, and (3) tip-down bunched /r/.

![Fig. 2. Spectrograms of spoken by Subject 4EJ (top) and Subject 6FM (bottom).](image)

Each of the articulatory configurations for /r/ gives rise to a distinctive characteristic of /r/ -- extremely low F3 frequency that is often close to F2 (c.f. [7]). The range of values reported in the literature for the first three formants is approximately 250-550 Hz for F1, 900-1500 Hz for F2, and 1300-1950 Hz for F3 [1,4,7]. Two examples of intervocalic /r/ are shown spectrographically in Fig. 2. While the patterns for F1, F2 and F3 are very similar in both spectrograms, differences can be seen in the higher formants, particularly in F5. For the utterance on the top, F4 and F5 lower by about 200 Hz and 500 Hz, respectively, during the /r/. However, for the utterance on the bottom, F4 and F5 stay fairly flat between the /r/ and the adjacent vowels.
In this paper, we address this issue of the difference in the formant patterns seen above F3. More specifically, we use the simple tube model for /r/ developed in [8] to provide an explanation for the pattern of F3, F4 and F5 shown in the spectrograms of Fig. 2. We hope to show that F5 is a good acoustic index of differences in tongue shape for /r/.

2. STIMULI

Three sets of data are used in this study. One set of data consists of acoustic and articulatory data simultaneously collected from 2 speakers as part of a larger study [5]. An electro-magnetic midsaggital articulometer (EMMA) system was used to track movements of six transducers attached to the tongue, lips and lower incisor. In this study, we examine the vertical and horizontal position of the three transducers attached to the tongue back, tongue middle and tongue front at the lowest point of F3 occurring during the production of the word /warav/ spoken in the carrier phrase “Say _____ for me”. Fig. 3 shows the transducer positions for the /r/s produced by the two speakers whose spectrograms are shown in Fig. 1. Note that Subject 4EJ produced a tip-up retroflex /r/ with the tongue dorsum lowered while Subject 6FM produced a bunched and tip-down /r/.

Since we don’t have vocal tract areas for Subject 4EJ and Subject 6FM, we base our discussion of the simple tube model’s ability to explain the different formant patterns we see for these speakers on Magnetic Resonance Imaging (MRI) data reported in [6]. The MRI data used here was obtained from a native American English male who produced a sustained syllable /t/. Area functions were measured separately for the sublingual space (see Table 2) and for the continuous supralingual space running from the lips to the glottis (see Fig. 4). Note the large palatal constriction in Fig. 4 which corresponds to a midsaggital vocal tract profile as in Type 4 of Fig. 1.

3. SIMPLE TUBE MODEL

The simple tube model developed in earlier work [8] is shown in Fig. 5. It was derived from Stevens’ model for a retroflex /r/ [10]. The major difference between the models occurs in the region anterior to the palatal constriction. Stevens’ model assumes a relatively small supralingual front cavity with dimensions equal to the area of the lip opening. Because MRI data suggests a large front cavity and a considerable lip constriction, our model uses two tubes, a larger one representing the front cavity and a narrower one representing the lip constriction.

Another difference occurs in the pharyngeal region. In Stevens’ model, the pharyngeal constriction is represented as a perturbation rather than as a tight constriction. To account for the stronger constrictions seen in the MRI data of [8], as well as...
subject-to-subject differences in the degree of pharyngeal constriction seen in Fig. 1, we allow for the possibility of modeling the part of the vocal tract behind the palatal constriction as three separate tubes.

In [8], speaker-specific dimensions for the simple tube model were derived from the MRI data. These dimensions were tested using a vocal tract modeling program [11]. Results showed that the resulting formants were within 110 Hz of those estimated with the actual MRI-derived area functions. Dimensions for speaker MI for the simple tube model are given in Table 4.

Table 4: Areas and lengths for the simple tube model.

<table>
<thead>
<tr>
<th></th>
<th>Area (cm²)</th>
<th>Length (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip</td>
<td>A_l = 1.62</td>
<td>L_l = 1.50</td>
</tr>
<tr>
<td>Front</td>
<td>A_f = 4.47</td>
<td>L_f = 1.50</td>
</tr>
<tr>
<td>Front +</td>
<td>A_f' = 3.05</td>
<td>L_f' = 2.70</td>
</tr>
<tr>
<td>Sublingual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatal Const.</td>
<td>A_p = 0.87</td>
<td>L_p = 3.00</td>
</tr>
<tr>
<td>Mid</td>
<td>A_m = 3.70</td>
<td>L_m = 4.20</td>
</tr>
<tr>
<td>Pharynx Const.</td>
<td>A_p = 1.93</td>
<td>L_p = 4.20</td>
</tr>
<tr>
<td>Back</td>
<td>A_b = 3.56</td>
<td>L_b = 3.60</td>
</tr>
<tr>
<td>Mid + Pharynx</td>
<td>A_b' = 3.25</td>
<td>L_b' = 12.0</td>
</tr>
<tr>
<td>+ Back</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. FORMANT CALCULATIONS BASED ON SIMPLE TUBE MODEL

Using the simple tube model together with the MRI derived areas for the model listed in Table 4, we were able to show that F3 results primarily from the cavities anterior to the palatal constriction which includes the sublingual space. In fact, we found that one function of the sublingual cavity is to increase the volume of the oral cavity between the palatal constriction and the lip constriction, thereby lowering the frequency of F3. A reasonable estimate of F3 was obtained by decoupling the front part of the vocal tract from the palatal constriction. Treating the oral cavity and the sublingual space as one large volume, F3 was calculated by summing the admittance of the combined oral and sublingual cavity

\[(\omega = \frac{1}{\rho c}) + \frac{1}{A_p/c} + \frac{1}{A_c/c} + \frac{1}{A_m/c} + \frac{1}{A_b/c} + \frac{1}{A_{b'}}/c \] and determining the first zero crossing. (Note that we use \(c = 35000 \text{ cm/sec}\).

![Fig. 5. Simple tube model developed in [8]](image)

In [8], we also show that F1, F2 and F4 result from the cavities behind the palatal constriction. If the pharyngeal constriction is tight enough, F4 is the half-wavelength resonance of the mid cavity. If, however, the pharyngeal constriction is weak so that a mid cavity is not formed, then F4 is the second half-wavelength resonance of the large back cavity which extends from the glottis to the palatal constriction.

F5 will be the half-wavelength resonance of the next longest cavity. As we will show below, when the sublingual space is large, the next longest cavity will be the combined front cavity and sublingual space occurring between the palatal constriction and the lip constriction. In our data, the largest sublingual space occurs when the tongue tip is raised and the tongue dorsum is lowered as in the tip up retroflex /r/.

4.1 Model for Subject 4EJ

The model for Subject 4EJ will differ from the models for MI and Subject 6FM in at least two important ways. Reasoning that Subject 4EJ’s configuration resembles Type 7 of Fig. 1, we assume that Subject 4EJ forms the palatal constriction with only the tongue tip. Thus, the palatal constriction will be much shorter than the 3 cm seen in Table 4 for MI. Although we don’t have areas for Subject 4EJ, we can take advantage of the articulatory data for the Tamil tip up /r/ [9] which also resembles Type 7 of Fig. 1. Using a value for \(L_{pc}\) of 1 cm, a reasonable and conservative estimate for \(L_b\) is 2.5 cm, a value that is 1 cm longer than that specified in Table 4.

The other difference relates to the length of the sublingual cavity. Given that the tongue tip placement for Subject 4EJ is similar to that for the Tamil subject, the sublingual space should be comparable in these subjects and longer than what we observe for MI (see Table 2). Thus, we assume a value of 2.8 cm for \(L_{cs}\). This gives an effective length of the front cavity that is \(L_f = L_{pc} = 5.3 \text{ cm}\).

Taking this value of \(L_{pc}\) together with the values of \(A_p\), \(A_{b'}\) and \(A_{b''}\) listed in Table 4 for MI’s syllabic /r/ results in an estimated value of F3 of 1591 Hz. This estimate from the model is consistent with the values measured from this speaker’s acoustic data (see Table 1).

As stated above, F4 can be modeled in two ways. Because the transducer placement shown in Fig. 3 does not inform us about the degree of pharyngeal constriction, we will estimate F4 both ways. If the pharyngeal constriction is narrow, then F4 is the first half-wavelength resonance of the mid cavity. Again, since the palatal constriction is only about 1 cm long (instead of the 3 cm listed in Table 4), we assume that the mid cavity will be about 1 cm longer than the 4.5 cm listed in Table 4. Thus, the resulting frequency for F4 is 3182 Hz (\(c/2L_{cs}\)).

Alternatively, assuming a weak pharyngeal constriction, F4 will be the second half-wavelength resonance of a long back cavity. Taking into account the short palatal constriction, \(L_b\) will be around 5.5 cm. This value results in an estimate of the F4 frequency of 2692 Hz (\(c/2L_{b'}\)).

Based on the F4 frequencies measured from Subject 4EJ’s acoustic data, the model with the pharyngeal constriction yields a better estimate of F4. Thus, we hypothesize that Subject 4EJ has a narrow pharyngeal constriction.

Given the estimated lengths of all of the cavities in Subject 4EJ’s production of /r/, F5 is the second resonance of that part of the vocal tract that is anterior to the palatal constriction. That is, F5 is the first half-wavelength resonance of the large volume cavity between the lip constriction and the palatal constriction. With \(L_{pc}\) estimated at 5.3 cm, the frequency of F5 is 3300 Hz (\(c/2L_{pc}\)). This frequency of F5 is comparable to the values of F5.
measured from this subject’s actual acoustic production (see Table 1).

Thus, the simple tube model with the dimensions estimated from various articulatory data gives reasonable estimates of F3, F4 and F5. This model can also explain the amplitude characteristics of these formants. Note that Subject 4EJ has a weak amplitude for F5. Note also that in the model, the sublingual space acts as a side branch and accordingly will result in an antiresonance. If we assume a length of 2.8 cm for \( L_s \), the first zero of this quarter-wavelength tube will be around 3125 Hz (c/4L_s) which is close in frequency to F4 and F5. Thus, it is not surprising that F5 is so weak. Ordinarily, the closeness of F4 and F5 would suggest a boosted amplitude for both formants.

4.2 Model for Subject 6FM
As opposed to the tip-up retroflex configuration of Subject 4EJ, Subject 6FM’s tongue configuration appears to be a classic tip-down bunched /rt/. The closest configuration appears to be that of M1’s syllabic /rt/. Therefore we use the dimensions specified in Table 4. In this case, the Helmholtz shape of the front part of the vocal tract results in an estimate for F3 of 1699 Hz. This value is in the range of values measured from Subject 6FM’s acoustic data.

For Subject 6FM, F5 will not result from the half-wavelength cavity between the lip constriction and the palatal constriction since this cavity is relatively short (L\(_f\)=2.7 cm). If we assume that the pharyngeal constriction is tight enough so that it is modeled as a separate cavity, then F4 comes from the mid cavity and F5 will come from the pharyngeal constriction. Using the values of L\(_m\)=4.5 and L\(_s\)=4.2, F4 will be around 3900 Hz (c/2L\(_m\)) and F5 will be around 4167 Hz (c/2L\(_s\)). These values are consistent with those observed in Fig. 2 and listed in Table 1. Again, the data suggests that Subject 6FM has a tight pharyngeal constriction.

5. DISCUSSION AND CONCLUSION
In previous work we have shown that the simple tube model gives reasonable estimates of the F1, F2 and F3 values seen across several variants of /rt/. In this paper, we extend this work and show that the model also gives reasonable estimates of F4 and F5.

In the case of the tip-up retroflex /rt/, the lowered tongue dorsum presumably results in an increase to the length of the cavity behind the palatal constriction, resulting in a lowered F4. Additionally, the longer sublingual space and shorter palatal constriction will lengthen the front cavity, resulting in a very low F5.

In the case of the bunched /rt/ with the tongue tip raised or lowered, the longer palatal constriction presumably results in a rather short front cavity (first half-wavelength resonance will be above 5000 Hz). Thus, both F4 and F5 will result from cavities behind the palatal constriction. This back part of the vocal tract will include a mid cavity due to a tight pharyngeal constriction. Contrary to the pattern for Subject 4EJ, this results in no lowering of F4 and F5. Critical to the simple tube model is the “Helmholtz” shape of the cavities anterior to the palatal constriction. Note that a quarter wavelength shape for this part of the vocal tract cannot explain the relationship we see between F3 and F5 for these examples of American English /rt/.