THE ROLE OF F2 AND F3 IN THE PERCEPTION OF RHOTICITY: EVIDENCE FROM LISTENING EXPERIMENTS

Barry Heselwood & Leendert Plug

University of Leeds, UK
b.c.heselwood@leeds.ac.uk; l.plug@leeds.ac.uk

ABSTRACT

This paper reports on two listening experiments which explore the relative contributions of F2 and F3 in the perception of rhoticity. Experiment 1 tests the hypothesis that a low-frequency F3 is a crucial spectral component for a signal to be perceived as rhotic. Its results suggest that the removal of F3 may in fact strengthen the rhoticity percept. Experiment 2 manipulates the relative amplitudes of F2 and F3 in a rhotic signal. Its results suggest that F2 and F3 attenuation have opposite effects on rhoticity judgments, with F3 attenuation making the signal sound more rhotic. The paper proposes an account that makes crucial reference to the auditory integration hypothesis.

Keywords: rhoticity, formants, perception, auditory integration, English

1. INTRODUCTION

Discussion of the acoustics of rhoticity (in this paper, the percept associated with post-alveolar or retroflex approximation) tends to focus on the importance of a low-frequency F3 [3, 4, 6]. Early studies using pattern playback to synthesize sequences with English /l/ and /r/ [9, 10] found that the value of F3 is important in distinguishing /l/ from /r/: 'the third-formant onset of /r/ needs to be lower in frequency, fairly close to the second-formant onset' [10: 34]. Since these studies, the view that a low-frequency F3 contributes crucially to the perception of rhoticity has remained largely unchallenged. Stevens [11] has argued that what is usually interpreted as a drop in F3 frequency at the onset of /r/ production is in fact the emergence of an additional formant, F_R, 'in the frequency range normally occupied by F2', along with a drop in amplitude of F3 proper [11: 540-541]. Still, it has remained customary to refer to a low-frequency third formant as the principal acoustic correlate of rhoticity, and some authors assume a direct correlation between the third formant value and the degree of perceived rhoticity: 'the lower the F3, the greater the degree of rhoticity' [8: 149].

To our knowledge, the latter assumption has not been proven experimentally, and the question remains whether it is the presence of a low-frequency F3/F_R *per se* that is responsible for the rhoticity percept. Of the studies cited above, several highlight the closeness of the third formant to F2 [9, 10, 11]. According to the auditory integration hypothesis, two acoustic formants form a single perceptual component if they are within 3.5 Bark of each other [1, 2, 5]. If this is correct, listeners are unlikely to be able to separate out the third acoustic formant in rhotic productions; instead, a single perceptually integrated formant in the F2 frequency range will be perceived [5: 167].

These considerations warrant experimental research to establish the relative contributions of the third formant (henceforth 'F3') and F2 to the perception of rhoticity. This paper presents results of some first steps in this direction, in the form of two listening experiments. Experiment 1 aims to establish whether a low-frequency F3 is a necessary or at least useful spectral component if a signal is to be perceived as rhotic. Experiment 2 follows this up to explore the relative contributions of F2 and F3 to rhoticity perception.

2. EXPERIMENT 1

2.1. Aim and hypothesis

The aim of Experiment 1 was to establish what effect the removal of F3 from a rhotic signal has on the perceived degree of rhoticity. If a low-frequency F3 is a crucial acoustic correlate of rhoticity, the prediction must be that the removal of F3 will result in a decrease in the degree of perceived rhoticity.

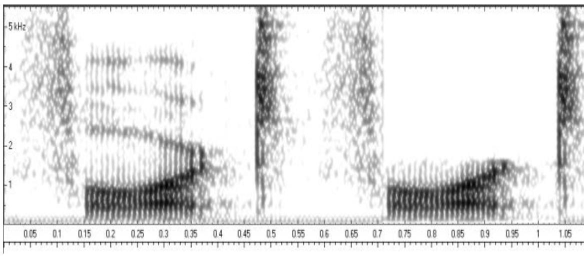
2.2. Method

2.2.1. Stimuli

Rhotic tokens of the words *fort, stars* and *hurt* were selected from a wordlist recording of a 79-year-old male speaker of English from Accrington, Lancashire, a residually rhotic area in north-west England. The three tokens were lowpass filtered in

Sensimetrics' *SpeechStation 2* to remove all acoustic energy above F2 during the vocalic portion. Figure 1 shows a spectrogram of the unfiltered and filtered version of *fort*. For *hurt* the filtering was applied also to the initial /h/ because it contains spectral prominences continuous with the following vocalic formants which might cue rhoticity. For each token, a stimulus was created consisting of the unfiltered version followed by the filtered one (*fort*, *stars*), or vice versa (*hurt*).

Figure 1: Spectrogram of the stimulus for *fort*, consisting of the original rhotic form (left) and the lowpass-filtered version (right).



2.2.2. Participants and procedure

Forty academic phoneticians participated in the experiment. They were sent the three stimuli as separate sound files by email. All were instructed to listen to each stimulus twice only and to judge which of the two forms sounded more rhotic, with 'both sound equally rhotic' as a third option. The participants were not told how the stimuli were prepared, and they were explicity told not to carry out instrumental analysis. No restrictions were placed on listening conditions.

2.3. Results

The results are summarized in Table 1. For all three tokens, a clear majority of listeners judged the filtered version to be at least as rhotic as the unfiltered one. In the case of *fort*, 40% judged the filtered version to sound as rhotic as the unfiltered version, and 45% judged the filtered version to be more rhotic. For *stars*, 80% heard the filtered version as more rhotic than the unfiltered one, and for *hurt*, 73% reported the same. Despite the informal set-up of the experiment, these results clearly suggest that contrary to the starting hypothesis, the removal of F3 from a rhotic signal does not necessarily result in a decrease in the degree of perceived rhoticity; in fact, for most listeners it results in an increase.

Table 1: Rhoticity judgments of rhotic *fort*, *start* and *hurt*, and their lowpass-filtered equivalents.

	Degree of rhoticity of filtered version relative to unfiltered		
	less	equal	more
fort	6 (15%)	16 (40%)	18 (45%)
stars	4 (10%)	4 (10%)	32 (80%)
hurt	5 (13%)	6 (15%)	29 (73%)

3. EXPERIMENT 2

3.1. Aim and hypothesis

The results of Experiment 1 clearly warrant more carefully controlled experimental work to establish the relative contributions of F2 and F3 to rhoticity perception. The aim of Experiment 2 was to establish what effect the alternate attenuation of F2 and F3 in a rhotic signal has on the degree of perceived rhoticity. If a low-frequency F3 is a crucial acoustic correlate of rhoticity, attenuation of F3 should result in a decrease in the degree of perceived rhoticity. Attenuation of F2, on the other hand, might be expected to heighten the percept of rhoticity, as F3 becomes a more prominent spectral component. In fact, the results of Experiment 1 suggest that the findings may well be the other way around.

3.2. Method

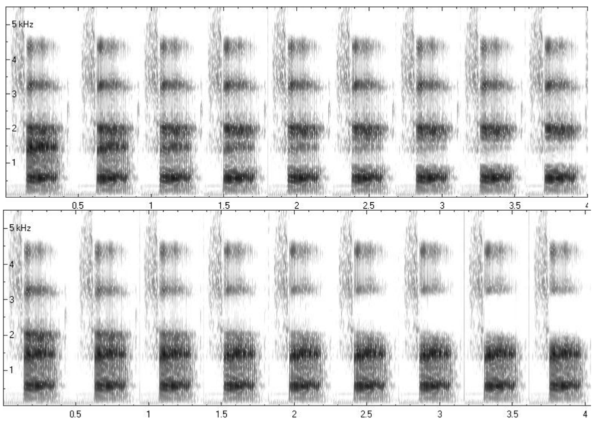
3.2.1. Stimuli

A rhotic token of the word *fir* was selected from the same recording as that used for Experiment 1. The token was perceived by two independent academic phoneticians to have a fully rhotic vocalic portion. The portion shows minimal formant movement, with F2 at 1444Hz and F3 at 1954Hz at the midpoint. F2 constitutes the dominant formant peak in the spectrum, and the difference in amplitude between F2 and F3 is 4.7dB. The token was filtered in *SpeechStation 2* such that either F2 or F3 was attenuated in eight 3dB steps. This results in 16 additional forms with varying F2-F3 amplitude relations. Figure 2 shows spectrograms of the resulting continua. Each filtered form was combined with the original to create 32 stimuli consisting of the original rhotic token followed or preceded by a filtered version with a weakened F2 or F3.

3.2.2. Participants and procedure

Forty-four phonetics students participated in the experiment. All were familiar with the concept of rhoticity and reported normal hearing. All did the experiment using a computer and headphones in a quiet room. Following familiarization, all heard the 32 stimuli in random order, with the addition of 6 filler pairs consisting of two different word forms which were clearly rhotic or non-rhotic. Each pair was played three times, after which participants judged which member sounded more rhotic by clicking on a button on their computer screen. They were also given the response option ‘neither sounds rhotic’, which was appropriate for one of the filler pairs. They were not given the response option ‘equally rhotic’, and it was explained that if they heard any rhoticity, they had to decide on the relative degree in the two forms they heard.

Figure 2: Spectrograms of the continua resulting from stepwise reduction of F2 (top) and F3 (bottom) in rhotic *fir*.

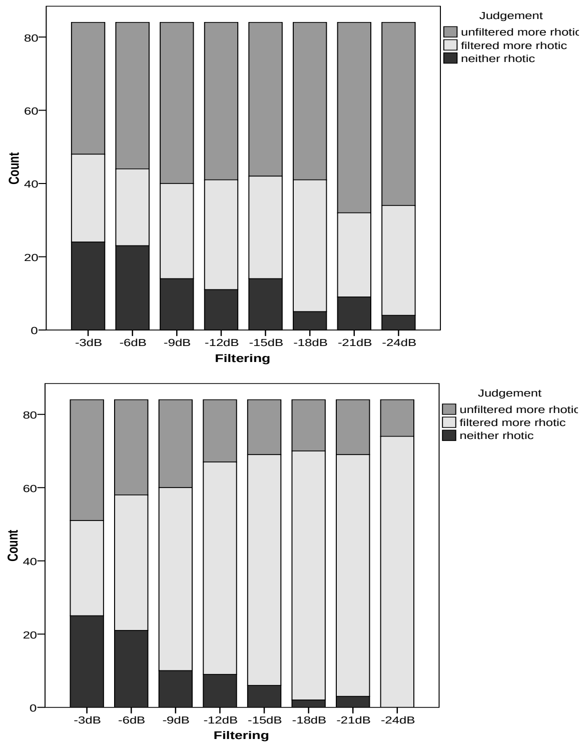


3.3. Results

Two participants were excluded due to poor performance on the filler pairs. The responses of the remaining 42 participants showed a high degree of conformity (ICC=0.92). The response ‘neither sounds rhotic’ was unexpectedly frequent, in particular for pairs containing minimally distinct tokens; it became less frequent the more extensive the formant attenuation. This suggests some participants may have used it to express that the two members of a pair sounded *equally* rhotic. Results are summarized in Figure 3.

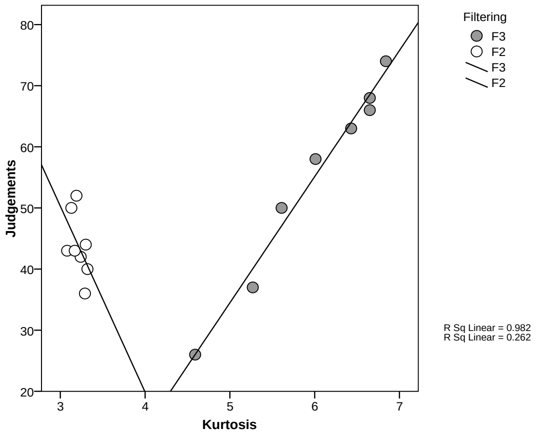
Figure 3 shows that as F2 is attenuated, the proportion of ‘unfiltered more rhotic’ responses increases, although not in neat steps. The highest proportion (52 responses, 62%) is observed at –21dB. The proportion of ‘filtered more rhotic’ responses fluctuates without showing a clear trend. The results for F3 attenuation clearly show that the lower the amplitude of F3, the more rhotic the filtered form sounds. Except for –21dB, the proportional responses show a neat stepwise pattern, with a maximum of 74 ‘filtered sounds more rhotic’ responses (88%) at –24dB. These results are consistent with those of Experiment 1: weakening the relative contribution of F3 to the overall spectral envelope results in a strengthening of the percept of rhoticity. Attenuating the relative contribution of F2 appears to have an effect in the opposite direction, albeit a less striking one.

Figure 3: Responses for ‘unfiltered more rhotic’ (dark grey), ‘filtered more rhotic’ (light grey) and ‘neither rhotic’ (black): F2 (top) and F3 attenuation (bottom).



Spectral analysis of the stimuli shows that rhoticity judgments are closely correlated with the kurtosis of the averaged spectrum during the vocalic portion in *fir*: that is, with the extent to which the spectrum is dominated by a single peak [7, 12]. Figure 4 shows that F3 attenuation results in an increase in kurtosis which is strongly correlated with the number of ‘filtered sounds more rhotic’ responses (R^2=0.96). For F2 attenuation, kurtosis values show a weak negative correlation with the number of responses, (R^2=0.26). The effect of filtering on kurtosis values requires further analysis, but these findings strongly suggest that listeners are sensitive to the prominence of a single dominant peak in the F2 frequency region.

Figure 4: Correlation between N responses for 'filtered more rhotic' (*y*-axis) and kurtosis of the average spectrum of the vocalic portion (*x*-axis), for F2 attenuation (white) and F3 attenuation (grey).



4. DISCUSSION

Our results suggest that the crucial auditory correlate of rhoticity is not a low-frequency F3 *per se*, but rather a single perceptual formant in the F2 region, which we might label F-rho (Fρ). The correlation of rhoticity judgments with spectral kurtosis suggests that the more Fρ dominates the auditory spectrum, the more robust is the perception of rhoticity. In this account, removal or attenuation of acoustic F3 has two effects: it shifts the frequency value of Fρ down towards acoustic F2 and away from higher spectral components, and it increases the dominance of F2 in the acoustic spectrum – and consequently the dominance of Fρ in the auditory spectrum. This is consistent with listeners' responses to F3 removal in Experiment 1, and to F3 attenuation in Experiment 2.

Our account makes crucial reference to the auditory integration hypothesis [1, 2, 5]. In the original production of *fir* in Experiment 2, acoustic F3 is within 3.5 Bark of both F2 and F4. Moreover, the formant that results from the auditory integration of F2 and F3, Fρ, is just outside the 3.5 Bark integration range from F4. When F3 is attenuated, the centre frequency of Fρ will remain outside of this range. F2 attenuation, on the other hand, shifts the centre frequency of Fρ towards that of acoustic F3 – that is, towards the auditory integration range around F4, lowering the perception of rhoticity. On this interpretation, the low amplitude and frequency of F3 in rhotic productions may be the result of its suppression by the vocal tract, in order to maintain F2 kurtosis and ensure that Fρ is free from interference from higher spectral components. Additional experiments are needed to examine the influence of F4 on rhoticity perception, and to test the predictions of the auditory integration hypothesis in a wider range of vowel contexts. As well as using natural speech stimuli from other speakers and other language varieties, future research should employ synthetic stimuli with carefully controlled frequencies and amplitudes to establish the thresholds of Fρ.

5. CONCLUSION

This paper has presented results from two listening experiments which suggest that the widely held assumption that a low-frequency F3 is a crucial acoustic and auditory correlate of rhoticity should be refined. The experiment results confirm that this assumption is valid only to the extent that the low frequency of F3 contributes to the dominance in the auditory spectrum of a single peak in the acoustic F2 frequency region. Paradoxically, its absence altogether increases this contribution, and therefore strengthens the percept of rhoticity.

6. REFERENCES

[1] Bladon, A. 1983. Two-formant models of vowel perception: shortcomings and enhancements. *Sp. Comm.* 2, 305-313.
[2] Chistovich, L.A., Lublinskaya, V.V. 1979. The 'center of gravity' effect in vowel spectra and critical distance between the formants. *Hearing Res.* 1, 185-195.
[3] Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S., Alwan, A. 2000. Acoustic modeling of American English /r/. *J. Acoust. Soc. Am.* 108, 343-356.
[4] Harrington, J. 2010. Acoustic phonetics. In Hardcastle, W., Laver, J., Gibbon, F. (eds.), *The Handbook of Phonetic Science* (2nd ed.). Oxford: Blackwell, 81-129.
[5] Hayward, K. 2000. *Experimental Phonetics*. London: Longman.
[6] Johnson, K. 2003. *Acoustic and Auditory Phonetics* (2nd ed.). Malden, Mass.: Blackwell.
[7] Jones, M., McDougall, K. 2009. The acoustic character of fricated /t/ in Australian English. *J. Int. Phon. Ass.* 39, 265-289.
[8] Ladefoged, P. 2003. *Phonetic Data Analysis*. Oxford: Blackwell.
[9] Lisker, L. 1957. Minimal cues for separating /w r l y/ in intervocalic position. *Word* 13, 256-267.
[10] O'Connor, J.D. Gerstman, L.J., Liberman, A.M., Delattre, P.C., Cooper, F.S. 1957. Acoustic cues for perception of initial /w j r l/ in English. *Word* 13, 25-43.
[11] Stevens, K. 1998. *Acoustic Phonetics*. Cambridge, Mass.: MIT Press.
[12] Tanner, K., Roy, N., Buder, E.H. 2005. Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy? *J. Voice* 19, 211-222.