RHOTICITY WITHOUT F3: LOWPASS FILTERING, F1-F2 RELATIONS AND THE PERCEPTION OF RHOTICITY IN 'NORTH-FORCE', 'START' AND 'NURSE' WORDS[1].

Barry Heselwood

Abstract

The results of two perception tests designed to shed light on the perception of rhoticity[2] are reported. In the first, rhotic tokens of the words *fort*, *stars* and *hurt* were played to forty phonetically trained listeners in two stimulus conditions: unfiltered and filtered. In the unfiltered condition F3 and all spectral components above it up to the Nyquist frequency of 5.5kHz were present. In the filtered condition all components above F2 had been removed by lowpass filtering. Contrary to received expectation, most listeners reported hearing stronger rhoticity in the filtered condition. This result is explained in terms of F1-F2 relations and the broad-band auditory integration hypothesis (Bladon 1983). In the second experiment, a non-rhotic token of the word *nurse* was played to twenty three phonetically trained listeners, a subset of the listeners for the first experiment, in an unfiltered and filtered condition. In the filtered condition F3 and all upper frequencies were once more removed. Again contrary to received expectation, and perhaps even more surprisingly, a clear majority of listeners heard the filtered token as rhotic. Taken together, the results of both experiments indicate that, far from inducing the perception of rhoticity, F3 may in fact have an inhibiting effect on it.

1. Introduction

Discussion of the acoustics of rhoticity tends to focus on the importance of a lowered F3 (e.g. Fujimura & Erickson, 1997: 81; Espy-Wilson, Boyce, Jackson, Narayanan & Alwan, 2000: 344; Johnson, 2003: 111). Ladefoged (2003: 149) sums up the generally accepted view when he says that '[v]ariations in the frequency of F3 indicate the degree of r-colouring: the lower the F3, the greater the degree of rhoticity'. While it is clear that F3 is consistently observed to descend through the frequency spectrum whenever /r/ follows a vowel, this fact does not necessarily entail that a low F3 is required to be present in order for rhoticity to be perceived. In her landmark study of 'r' variants, Lindau (1985: 165) regards a low F3 as 'a well-justified specification for the American English /r/' which is an approximant, but 'not a pervading property of rhotics', a class that includes non-approximants. Evidence is presented below from lowpass filtered tokens of vowel+[ɹ] and syllabic [ɹ̩] to show that F3 need not be present at all in the spectrum of the signal for listeners to perceive rhoticity in approximants. In the absence of F3, it appears that rhoticity is perceived in approximants when F2 is far enough away from F1 to escape auditory integration whilst remaining below a threshold of c.11.5 Bark. Furthermore, additional evidence is presented to show that the removal of F3 from a token that sounds non-rhotic can make it sound rhotic.

---

[1] This research was part-funded by British Academy grant number SG 45840. Sincere thanks are due to Leendert Plug and the anonymous reviewer for valuable observations and suggestions which have led to improvements in presentation and argumentation. Any remaining shortcomings are of course my own responsibility.

[2] In this paper the term 'rhoticity' refers to the perceived presence of non-pre-vocalic constrictive approximant realisations of /r/ of the [ɹ] type.

The concept of auditory integration concerns the fusion in perception of two or more simultaneous spectral components such as acoustic formants to form a single perceptual component (Chistovich & Lublinskaya, 1979). By converting formant frequencies to the psychoacoustic Bark scale[3] we can see if two formants are close enough to be fused into a single perceptual formant. It is claimed that this occurs when two formants are within 3.5 Bark of each other (Bladon, 1983: 311-13; Hayward, 2000: 154-6). Bark values for F1, F2 etc. are given hereafter as Z1, Z2 etc.; a perceptually fused formant with input of F1 and F2 is given as Z_1*, and with input of F2 and F3 as Z_2*.

The rhotic speech used in this study was recorded in Accrington, Lancashire, a residually rhotic area in the north-west of England surrounded by non-rhotic accents (Wells, 1982: 368). Recordings were made on a Marantz PMD671 digital recorder using a Shure SM48 mono unidirectional cardioid polar-pattern microphone in a quiet room of the speaker's home. The speaker was a 79-year-old male. The sampling rate was set at 24kHz with 16 bit resolution in wave file format. For formant analysis, the recordings were re-digitised at a sampling rate of 11025Hz. The non-rhotic token of *nurse* in figures 8, 9 and 12 was spoken by a 20-year old male from Accrington. The recording was made onto a computer in the Phonetics Laboratory at the University of Leeds.

Neither of the speakers recorded made a distinction between NORTH and FORCE words although membership of these lexical sets is not quite the same as for RP English. For example, *door, floor* are FORCE words in RP but CURE words in Accrington; *war, warn* are NORTH words in RP but tend to have a vowel more like the vowel in START in Accrington, at least amongst elderly speakers. NURSE words are kept distinct from SQUARE words with the same incidence as in RP.

2. Rhoticity perception thresholds

Some work on establishing psychoacoustic thresholds for the perception of rhoticity has been carried out by Heselwood, Plug & Tickle (in press). They report that after the front [ɛ] vowel thirty of their thirty two listening subjects reported hearing a [ɹ] sound when the value of Z_2* fell to 12.7 Bark. The other two subjects reported this effect at 12.5 Bark, i.e. a little later in the time course of the signal. After the back vowel [ɔ], in which Z2 and Z3 begin around 8.7 Bark apart, rhoticity was perceived by twenty eight listeners when they had approached each other to within 4.7 Bark and by the remaining four listeners when they were 3.9 Bark apart.

These threshold values are broadly in agreement with results reported by Yaeger-Dror, Kendall, Foulkes, Watt, Harrison, Kavanagh & Oddie (2009) using a larger number of listeners.

In this section, a rhotic token of a word from each of the three lexical sets NORTH-FORCE, START and NURSE spoken by the 79-year-old Accrington speaker, and a non-rhotic token of a word from the NURSE set spoken by the 20-year-old speaker, are analysed acoustically and psychoacoustically. The perceptual descriptions are based on the author's impressions of listening to them. The Hz and Bark values for all the tokens to be discussed are presented in the Appendix. Section 3 then reports results from two experiments using lowpass filtered versions of these tokens in which F3 has been removed.

---

[3] The equation for converting Hz values to Bark values is $Z = ((26.81*f)/(1960+f)) - 0.53$ (Traunmüller, 1990) where Z is a Bark value and *f* is a Hz value.

2.1. NORTH-FORCE words

During the vowel part of the vowel+/r/ sequence in words belonging to the lexical sets NORTH and FORCE, F2 is very close to F1 and quite distant from F3. The token of the word *fort* in figure 1 spoken by the 79-year-old speaker provides the example to be discussed. Towards the end of the sequence, F2 has moved away from F1 and F3 has descended so that F2 and F3 are now close together. This dynamic correlates with a shift in perception from the vowel to the realisation of the following /r/. Figure 2 presents the changing formant proximities in Z-transformed values, modelling the perception of rhoticity as an increase in Z1-Z2 distance and a concomitant decrease in Z2-Z3 distance.

Figure 1. Spectrogram of rhotic *fort* showing the four sampling points for measuring the Z1-Z2 and Z2-Z3 distances shown in figure 2.

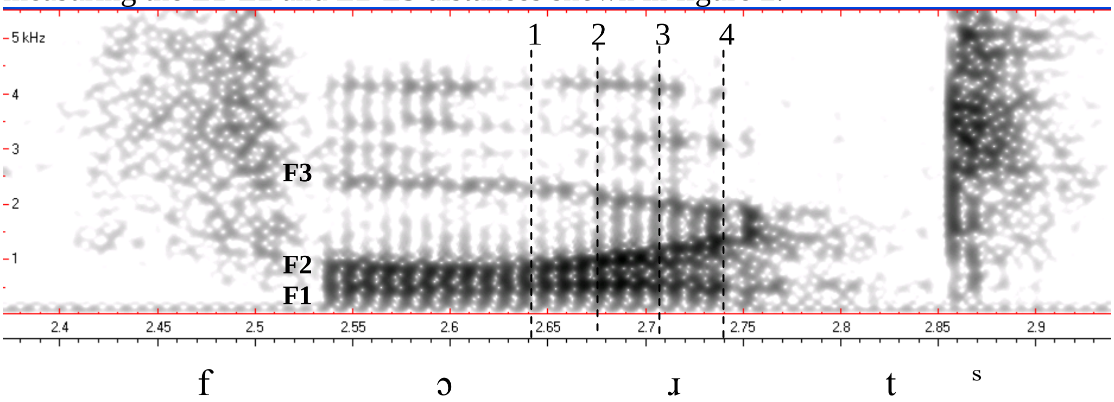
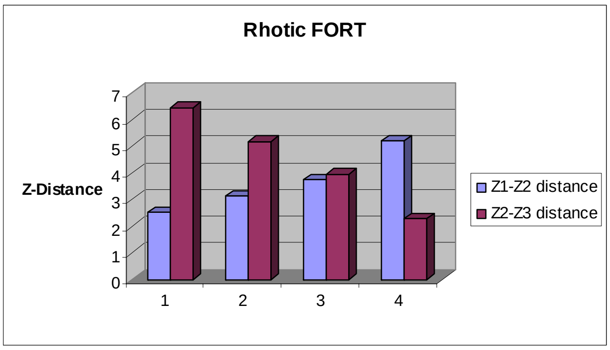


Figure 2. Z1-Z2 and Z2-Z3 distances in Bark at the measurement points indicated in figure 1.



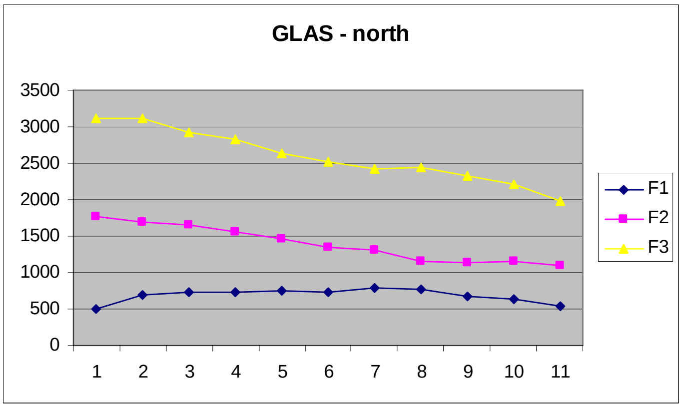
At points 1 and 2 in figures 1 and 2, the Z1-Z2 distance is less than the auditory integration value of 3.5 Bark while the Z2-Z3 distance is greater than the 4.7 Bark rhoticity perception threshold suggested by Heselwood et al (in press). It is therefore to be expected that up to and including point 2 no [ɹ] will be perceived and to the author the impression is of a [ɔ] monophthong[4]. At point 3 the Z1-Z2 distance has increased and the Z2-Z3 distance has decreased to below 4.7 Bark but is still slightly greater than the Z1-Z2 distance and the vowel changes to a schwa-like quality. By the end of the sequence, at point 4, the Z1-Z2 distance has surpassed the Z2-Z3 distance

[4] All perceptual judgements reported in this paper concerning where rhoticity begins are those of the author and were made before formant measurements were taken. Further listening experiments using gating are planned to see the extent of inter-listener agreement.

which is now well below 4.7 Bark. It is between points 3 and 4 that rhoticity begins to be perceived.

Evidence that a low F3 is not always sufficient to trigger perception of [ɹ] comes from the token represented in figure 3 in the form of formant tracks (from Heselwood et al., in press). F3 descends to 1983Hz in this production of the word *north* by a speaker from Glasgow but there is no perception of rhoticity. As at point 3 in figures 1 and 2, Z1-Z2 and Z2-Z3 distances are about equal at the last measurement point in figure 3, with values of 3.74 Bark and 3.78 Bark respectively. Despite the Z2-Z3 distance being below the 4.7 Bark threshold, the fact that Z2 is no further from Z1 than it is from Z3 may be the reason why there is perception of a schwa offglide but no perception of [ɹ]. The role of Z1-Z2 distance is further investigated in section 3 below using lowpass filtering to remove F3.

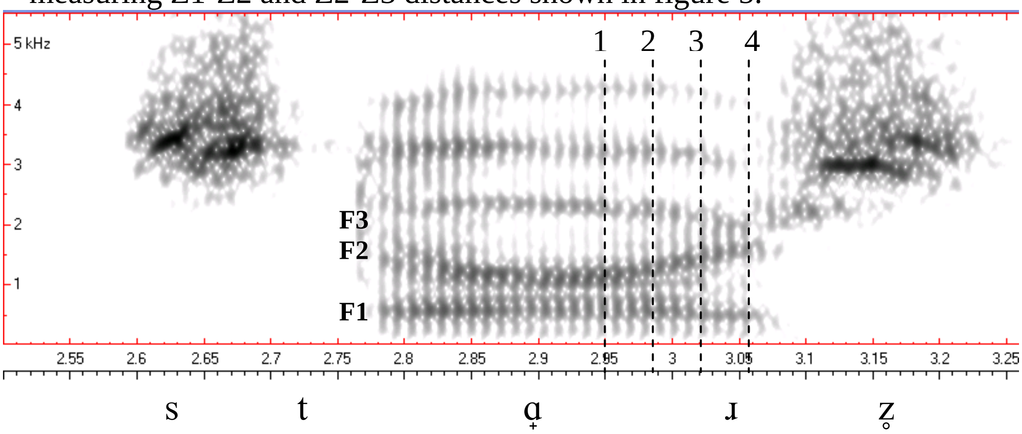
Figure 3. Low F3 in *north* but no perception of rhoticity (from Heselwood et al, 2009).



2.2. START words

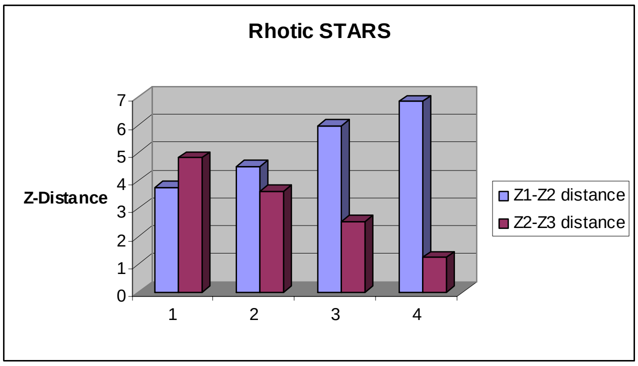
In the token of the word *stars* below, perception of rhoticity begins around point 3 on figure 4.

Figure 4. Spectrogram of rhotic *stars* showing the five sampling points for measuring Z1-Z2 and Z2-Z3 distances shown in figure 5.



At point 1 in figures 4 and 5, the Z1-Z2 distance is 3.76 Bark and although it is greater than the 3.5 integration threshold it is less than the Z2-Z3 distance of 4.81 Bark. There is no perception of rhoticity at this point and the vowel sounds monophthongal [ɑ̟].

Figure 5. Z1-Z2 and Z2-Z3 distances at the points indicated in figure 4.



At point 2, the Z1-Z2 distance is further outside the 3.5 Bark integration band at 4.52 Bark, and the Z2-Z3 distance is now reduced to 3.62, close to the integration band value. Perceptually, there is a slight offglide towards a schwa quality. At points 3 and 4, the Z1-Z2 distance has increased further while Z2-Z3 distance has dereased to well below the 3.5 Bark threshold. Rhoticity increases in perceptual strength from point 3 to point 4, correlating with the presumed increase in the amplitude of the articulatory gestures responsible for realising the post-vocalic /r/.

2.3. NURSE words

There is very little formant movement in the token of the word *hurt* (see figure 6), and rhoticity is perceived throughout the whole of the syllable nucleus. Rather than a sequence of vowel+/r/, the phonetic facts support an analysis in which /r/ is syllabic, i.e. [ɹ̩].

Figure 6. Spectrogram of rhotic *hurt* showing the sampling points for the measurements given in figure 7.

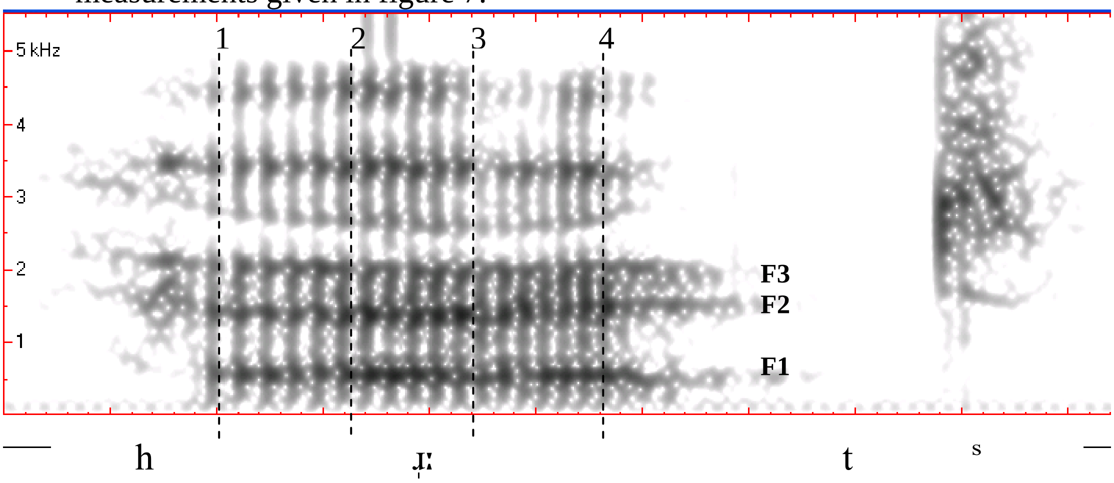
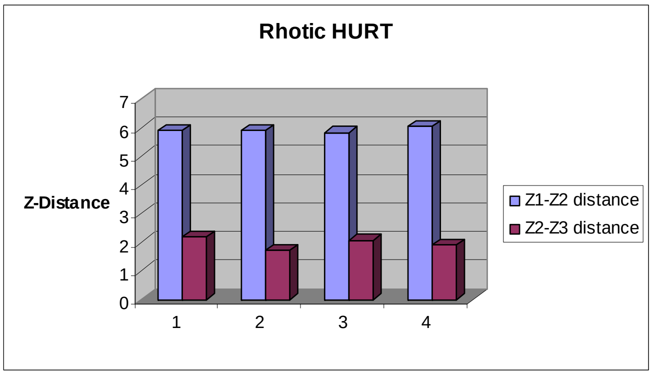


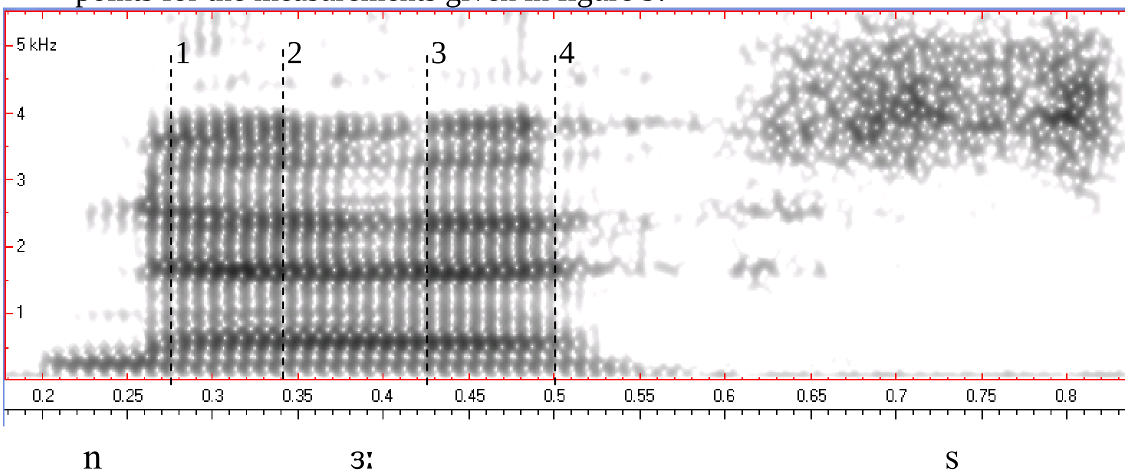
Figure 7 shows that the Z1-Z2 distance remains well outside the 3.5 Bark integration band while the Z2-Z3 distance remains well within it. Hayward (2000: 167) makes the same observation for American English [ɚ] in e.g. *bird*, noting that it is ‘highly unusual in showing a strong peak, resulting from integration of F2 and F3, in the middle of the range of possible F2 values’.

Figure 7. Z1-Z2 and Z2-Z3 distances at the points indicated in figure 6.



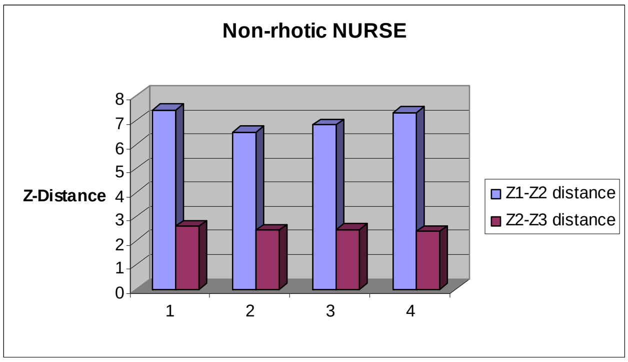
The Z-distances throughout *hurt* are similar to those at measurement point 4 in *fort* and *stars* in figures 2 and 5 respectively, the points at which rhoticity is most strongly perceived. The interesting question is whether it is the integration of Z2-Z3 that triggers perception of rhoticity as Hayward suggests, or whether it is the escape of Z2 into a region of auditory space where it avoids integration with Z1. This question is explored in section 4.4 below. Meanwhile, it is interesting to look at the non-rhotic token of *nurse* in figures 8 and 9 produced by the 20-year-old Accrington speaker.

Figure 8. Spectrogram of a non-rhotic token of *nurse* showing the sampling points for the measurements given in figure 9.



The Z2-Z3 distance, as in the case of *hurt* above, remains well within 3.5 Bark but this token, contrary to what Hayward's interpretation of F2-F3 relations would predict, is not rhotic. The features that distinguish this non-rhotic *nurse* from the rhotic *hurt* appear to be the greater Z1-Z2 distance which exceeds 6 Bark all through *nurse* but barely reaches 6 Bark in *hurt*, and the fact that Z2 in *nurse* is always higher than 11 Bark, but always lower than 11 Bark in *hurt* except at the end where it touches 11.10 Bark. We will see later in section 4 the surprising effect of removing F3 from the *nurse* token which prompts an alternative explanation for why the token in figure 8 sounds non-rhotic.

Figure 9. Z1-Z2 and Z2-Z3 distances at the points indicated in figure 11.



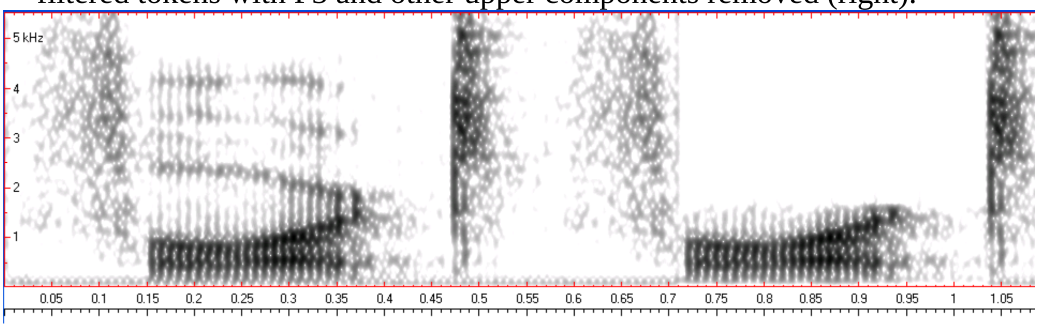
3.0. F3 and perception of rhoticity

If the presence of F3 is essential for perception of rhoticity, then the prediction would be that if F3 is removed from a rhotic token it will not be perceived as rhotic. To test this prediction, the first of the two perception experiments was carried out.

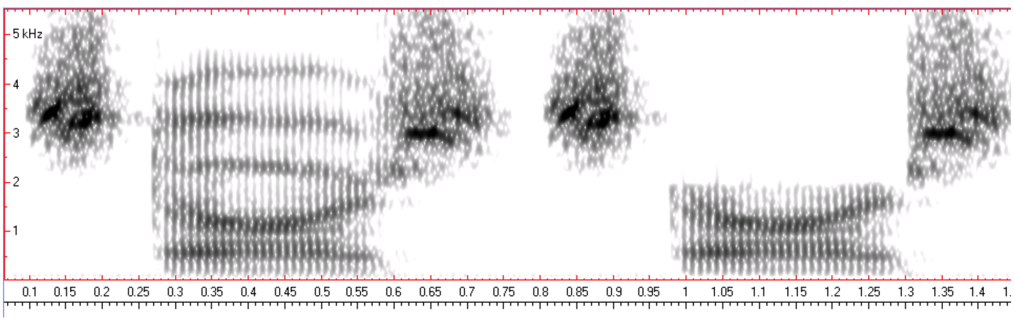
3.1. Stimuli

The rhotic tokens represented spectrographically in figures 1, 4 and 6 were lowpass filtered to remove all acoustic energy above F2 during the vocalic portion. In *hurt* the filtering was applied also to the initial /h/ because it contains spectral prominences continuous with the following vocalic formants and thus anticipates the formants in the syllabic [ɹ̩] which might cue rhoticity. Figures 10a-c show spectrograms of the un-filtered and filtered versions for comparison.

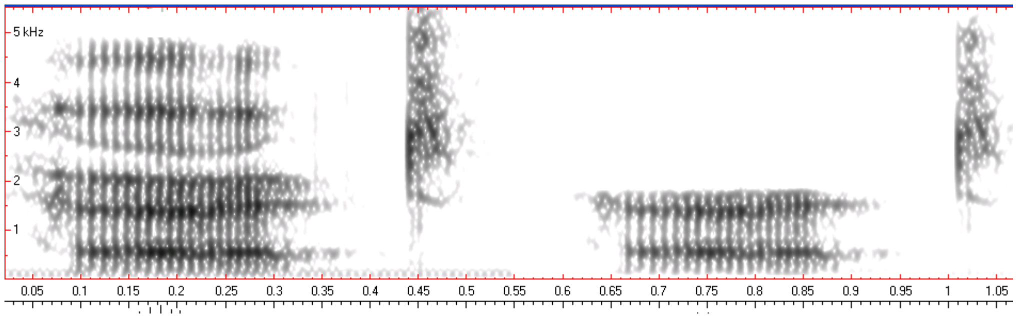
Figure 10. Spectrograms of the original unfiltered tokens (left) and the lowpass filtered tokens with F3 and other upper components removed (right).



a) *fort (unfiltered)* *fort (lowpass filtered, cut-of frequency = 1536Hz)*



b) *stars (unfiltered)* *stars (lowpass filtered, cut-off frequency = 1837Hz)*



c) *hurt (unfiltered)* *hurt (lowpass filtered, cut-off frequency = 1630Hz)*

The final consonant was excluded from the scope of the filter. Being alveolars, there would be very little left if the upper frequencies were removed such that word recognition might be affected.

For each of the three test-words, a sound file was created containing the unfiltered token and the filtered token. These constituted the stimuli for the experiment. In the files for *fort* and *stars* the filtered token was presented first; in the file for *hurt* the unfiltered token was presented first.

3.2. Procedure

Forty phonetically trained volunteer subjects were sent the sound files and the response sheet by email[5]. They were instructed to listen to each file twice only and to say for each token whether it sounded rhotic or not. They were not told that the files contained filtered versions of the unfiltered tokens. The instructions explained that it was their impressionistic judgement that was of research interest and they were asked not to do any instrumental analysis. Five more subjects did the task in the presence of the author. In the case of both tokens sounding rhotic, subjects were instructed to say whether the first token sounded more or less rhotic than the second, or if both tokens sounded equally rhotic. No restrictions were placed on listening conditions. Subjects were free to choose free-field listening or to listen through headphones or earbuds but were asked to report which they had done. Eighteen had listened in free-field conditions, sixteen said they had listened through headphones, and one through earbuds; for the remaining five listeners there is no record of listening conditions. No noticeable effects of listening conditions were evident other than that the use of headphones tended to make it less likely that a subject would report hearing both tokens as equally rhotic, suggesting that headphones may enable a decision to be made more confidently one way or the other. However, this is not a very strong tendency.

3.3. Results

The results of the rhoticity judgement task are presented in table 1 and figure 11. What is immediately striking is that for *stars* and *hurt* the filtered token, that is the token with F3 and all upper frequencies removed, was heard by a very clear majority of the listening subjects as more rhotic than the corresponding unfiltered token containing F3; this despite the possible presence of coarticulatory effects in the post-

[5] I would like to thank all those who took part in this research as listening subjects.

/r/ consonants that might cue rhoticity (Plug & Ogden, 2003), suggesting that if such cues were present, listeners were not much influenced by them.

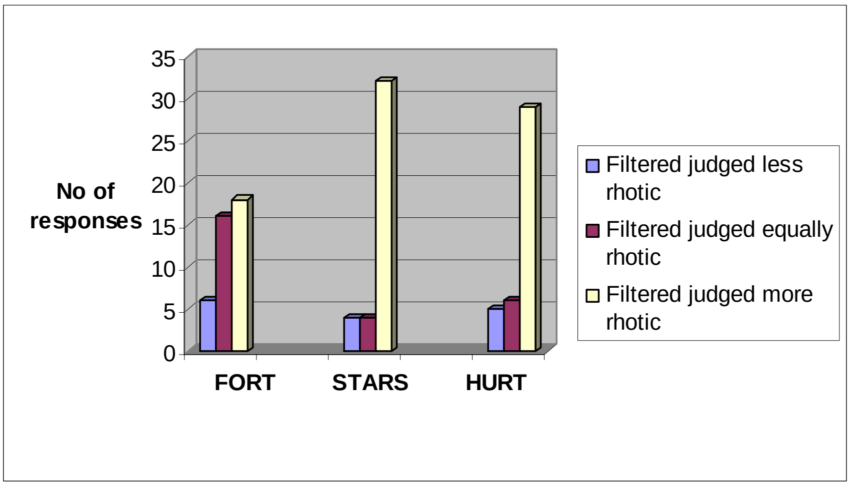
In the case of *fort* the position is not so clear but of the three response categories the one with the highest number of responses is the one reporting the filtered token as sounding more rhotic. In all three cases, only a small minority of listeners reported hearing the filtered token as less rhotic than the unfiltered token.

Table 1. Rhoticity judgements of unfiltered and filtered tokens of *fort, stars* and *hurt*.

N=40		Rhotic		Rhoticity of F compared to U		
		YES	NO	less	equal	more
fort	U	39 (97.5%)	1 (2.5%)	6 (15.0%)	16 (40.0%)	18 (45.0%)
	F	38 (95.0%)	2 (5.0%)			
stars	U	33 (82.5%)	7 (17.5%)	4 (10.0%)	4 (10.0%)	32 (80.0%)
	F	40 (100%)	0			
hurt	U	31 (77.5%)	9 (22.5%)	5 (12.5%)	6 (15.0%)	29 (72.5%)
	F	38 (95.0)	2 (5.0%)			

U = unfiltered token F = filtered token

Figure 11. Judgement of rhoticity of filtered tokens compared to unfiltered tokens.



3.4. Discussion of results

The results from the lowpass filtering of rhotic tokens carried out in this first experiment call seriously into question the traditional view that F3 has an important role in the perception of rhoticity. When F3 and all higher spectral content are removed, listeners mostly still report hearing a rhotic quality. Indeed, the indication from the results is that rhoticity is perceptually more robust in the *absence* of F3. All three filtered tokens were judged rhotic by all listening subjects except for two subjects in the case of *fort* and two in the case of *hurt* (one being the same subject in both cases), whereas more listeners judged all three unfiltered tokens to be non-rhotic (seven and nine for *stars* and *hurt* respectively, and one for *fort*). That F3 may even have an inhibiting effect on perception of rhoticity in *stars* and *hurt* is suggested by the fact that for these items a large majority of listeners (32/40 = 80.0% and 29/40 = 72.5% respectively) reported the filtered token as sounding more rhotic than the unfiltered token. Lindau's scepticism concerning the importance of a low F3 for perception of rhoticity in non-approximant rhotics, mentioned above, can on the basis of the results of this experiment be generalised also to approximant rhotics, at least in the context of central and back vowels. An F3 value was clearly not a necessary component of the acoustic signal for the listening subjects to perceive a non-prevocalic approximant [ɹ] in the tokens presented to them in this experiment.

3.4.1. F1-F2 relations

In the filtered tokens only two acoustic formants are present – F1 and F2. The suggestion is explored in this section that perception of rhoticity in the light of the above results must be a function of the dynamics of the F1-F2 relationship through the course of the vocalic part of the syllable.

3.4.1.1. 'Fort'

F1 and F2 begin the vowel+/r/ sequence in the filtered token of *fort* very close together, as is expected for the mid-open back vowel [ɔ]. Up to the point where F2 begins to rise, F1 averages 450Hz and F2 775Hz. These translate into auditory terms as Z1 = 4.48 Bark and Z2 = 7.07 Bark. Z2 is therefore only 2.59 Bark above Z1, well within the 3.5 Bark integration band. Consequently, we can hypothesise that they constitute a single fused peak Z_1* in auditory space having a value of around 5.7 Bark. By the end of the sequence, F1 and F2 are considerably further apart with F1 = 476Hz and F2 = 1249Hz, meaning that Z1 = 4.71 Bark and Z2 = 9.90 Bark. The Z1-Z2 distance is now 5.19 Bark so Z2 will have escaped the integration skirts of Z1 to form its own peak of prominence in auditory space. As Z2 escapes from Z1, the conditions for the perception of rhoticity become established.

In non-rhotic monophthongal productions of NORTH-FORCE words, and in productions of THOUGHT words, F1 and F2 remain close throughout the vocalic part of the syllable such that the Z1-Z2 distance is always within 3.5 Bark. It is likely that where a schwa offglide is heard in NORTH-FORCE words, F2 rises such that the Z1-Z2 distance is close to the 3.5 Bark integration boundary.

3.4.1.2. 'Stars'

F2 begins relatively high in *stars* due to the positive transition from the preceding alveolar /t/ but descends towards the high F1 where it remains to form the vowel [ɑ] until it rises for the following [ɹ]. While F1 and F2 are in close proximity, their frequency values average 670Hz and 1120Hz respectively, equal to 6.30 Bark and 9.22 Bark. The Z1-Z2 distance is 2.92 Bark. At the end of the vowel+/r/ sequence the distance has increased past the integration threshold to 6.85 Bark (Z1 = 4.45 Bark, Z2 = 11.30 Bark). The change from [ɑ] to [ɹ] correlates with the escape of Z2 from Z1, as does the change from [ɔ] to [ɹ] in *fort*.

Non-rhotic monophthongal productions of START words, and productions of BATH and PALM words, show F1 and F2 remaining close together and their auditory correlates, Z1 and Z2, are always within 3.5 Bark. A slight rise in F2 and an increase in the Z1-Z2 distance to a value close to the 3.5 Bark boundary would be expected to cause perception of a schwa offglide.

3.4.1.3. 'Hurt'

In *hurt* F1 and F2 remain fairly stable throughout the syllable nucleus at values around 510Hz and 1450Hz respectively. These are equivalent to Z1 = 5.01 Bark and

Z2 = 10.87 Bark, giving a Bark separation value of 5.86 which is well in excess of the 3.5 Bark auditory integration band. We can thus hypothesise that the peaks in auditory space correlating with acoustic F1 and F2 are not fused into a single percept but constitute auditorily distinct peaks of prominence. In contrast to *fort* and *stars*, the Z2-Z3 distance is always less than the Z1-Z2 distance.

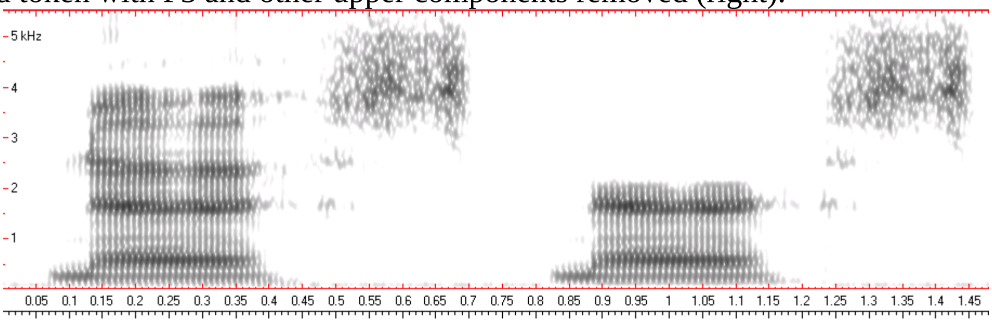
4.0. F3 and perception of non-rhoticity

The results from the first experiment suggest that tokens without F3 tend to sound more rhotic to most listeners. A second experiment was conducted to see if the removal of F3 from a non-rhotic token would have the effect of making that token sound rhotic.

4.1. Stimuli

In the second experiment the (non-rhotic) token of *nurse* shown in figure 8 was lowpass filtered to remove all spectral prominences above F2 from the nasal and the vowel. *Nurse* was chosen because the formant relations in words from the NURSE set tend to remain relatively stable throughout the syllable nucleus in naturally rhotic and non-rhotic productions as in figures 6 and 8. The stimuli for the experiment were the unfiltered and filtered versions of this word. Figure 12 shows the two versions for comparison.

Figure 12. Spectrogram of the original unfiltered token (left) and the lowpass filtered token with F3 and other upper components removed (right).



nurse (unfiltered) *nurse (lowpass filtered, cut-off frequency = 2006Hz)*

4.2. Procedure

The same procedure was used as for the first experiment but due to time constraints only twenty three of the original forty listening subjects took part. Twenty of the subjects were sent a sound file containing both versions of *nurse*, the unfiltered one coming first. Three more subjects did the task in the presence of the author. As before, subjects were not aware that one token was a filtered version of the other token. As was the case with the first experiment, listening conditions appear to have played no part in the exercise of the subjects' judgements. Of the seven who judged the filtered token to be non-rhotic, three listened with headphones, one with earbuds, and three in free-field conditions. Nine of the sixteen who reported hearing the filtered token as rhotic were wearing headphones, the other seven were not.

4.3. Results

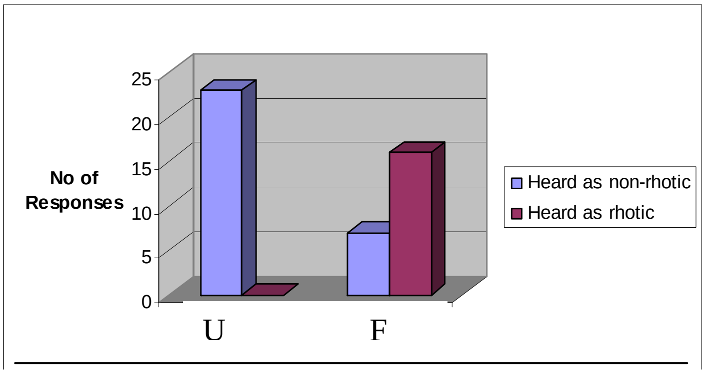
The results are shown in table 2 and figure 13. The unfiltered token was heard as non-rhotic by all listeners. Sixteen (69.6%) reported hearing the filtered token as rhotic. One of the seven reporting that it sounded non-rhotic commented nevertheless that it 'could be construed as rhotic'.

Table 2. Rhoticity judgements of unfiltered and filtered tokens of *nurse*.

N=23		RHOTIC	
		YES	NO
nurse	U	0	23 (100%)
	F	16 (69.6%)	7 (30.4%)

U = unfiltered token F = filtered token

Figure 13. Data in table 2 presented in bar-chart form.



4.4. Discussion of results

The situation with non-rhotic monophthongal productions of NURSE words is somewhat different than with non-rhotic NORTH-FORCE and START words. Instead of F1 and F2 remaining close within 3.5 Bark, they remain, as was seen in section 2.3 above, quite far apart such that the Z1-Z2 distance is well over the 3.5 Bark threshold as, of course, it also is in the case of rhotic productions. Recall that what appeared to distinguish non-rhotic *nurse* from rhotic *hurt* were the higher values for Z2 and for the Z1-Z2 distance (see section 2.3). However, the result of the second experiment indicates that what really distinguishes them is the stronger presence of F3 in non-rhotic *nurse*. Averaged over the vocalic portion of the syllable, the F3 peak is 4.4dB lower than F2 in non-rhotic *nurse*, but 6.5dB lower in rhotic *hurt* (see figure 15). The result further confirms that the presence of F3 is not a requirement for perception of rhoticity, that indeed it acts to supress perception of rhoticity, and that Z2 on its own is the crucial factor. From the listeners' responses to the filtered version of *nurse* we can predict that the 'classic' neutral vowel having F1 = 500Hz, F2 = 1500 and F3 = 2500Hz (Stevens & House, 1961: 308-310) will sound rhotic in the absence of F3. The resulting spectrum will have a Z2 value of 11.09 Bark and a Z1-Z2 distance of 6.17 Bark, almost identical in fact to the values at the end of the filtered *hurt* token in figure 10c. Future experimental work using synthetic speech stimuli could test this prediction.

Concerning the role of F3 in the perception of rhoticity in central and back vowel contexts, we are led by the results of the second experiment to the very opposite of received opinion. That is to say, we are led to state that the role of F3 is to prevent perception of rhoticity when F2 meets the conditions that would otherwise cause rhoticity to be perceived. In the unfiltered version of *nurse*, Z3 is well within 3.5 Bark of Z2 (see figure 9) and yet all listeners agree it is non-rhotic. If the auditory integration hypothesis is correct, then Z3 cannot be forming a separate peak in the auditory spectrum. Somehow it is suppressing the perception of rhoticity whilst being integrated with Z2 to form a fused Z_2*. Either it is pulling the value of Z_2* up beyond a rhoticity threshold on the Bark scale (vertical dotted line in figure 13a), or else it is broadening, and hence damping (Rosen & Howell, 1991: 192-3), the auditory peak to the point where it falls, on a loundness scale, below the sharpness needed to trigger perception of rhoticity (horizontal dotted line in figure 14b). Further investigation is needed to pursue this issue.

Figure 14. Two ways in which Z3 might prevent Z2 triggering perception of rhoticity whilst being auditorily integrated with it (see text for explanation).

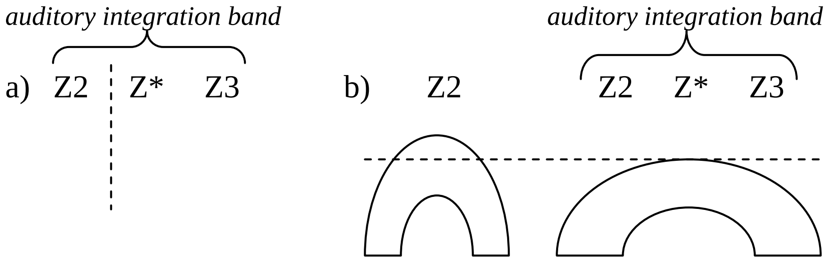
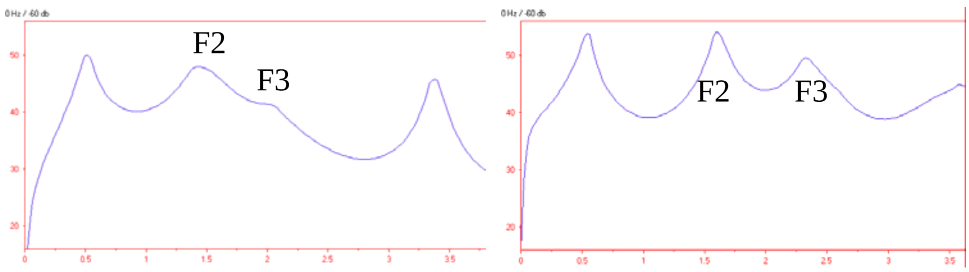


Figure 15. Averaged LPC spectra of unfiltered rhotic *hurt* (left) and unfiltered non-rhotic *nurse* (right). 128-point Hamming window.



The reason why Z3 prevents perception of rhoticity in the non-rhotic *nurse* but not in the rhotic *hurt* must be due to the much higher amplitude of the acoustic F3 in the former (see spectra in figure 14). In the latter it may simply not be strong enough to counter the rhoticity-inducing effect of Z2 in the majority of listeners. The low amplitude of other components above F2 no doubt also contributes to the auditory prominence of Z2 in *hurt*. In fact, it seems that the vocal tract configuration for [ɹ] has a lowpass filtering effect in its transfer function that is similar to the effect of the filters used in these experiments. To calculate the value of Z_2* of a given unfiltered token accurately, the relative amplitudes of F2 and F3 would need to be taken into account so as to derive a weighted average. The lower the frequency and amplitude of F3, the closer this value will be to Z2. The vocal tract appears in fact to be aiming to bring it as close as possible in frequency to Z2, and reduce its amplitude, in order to prevent it from inducing a non-rhotic percept. In other words, as the reviewer of this paper has suggested, what seems to be most important in inducing a rhotic percept is a steeply negative spectral balance between the middle of the F2 range and the region of the spectrum above it.

5.0. Conclusion

Perception of rhoticity appears to correlate for the great majority of listeners with the presence in auditory space of a distinct and relatively sharp peak in the region of 9.0-11.5 Bark, i.e. between about 1080-1595Hz. These values are in agreement with the F2 Bark values for postalveolar [ɹ] and retroflex [ɻ] given in Engstrand, Frid & Lindblom (2007: 178) although they encompass a slightly wider range than Engstrand et al. give. How low the peak can be and still be distinct depends on the value of F1. Because of the inverse relation between vowel height and F1 frequency, the peak will have to be higher the more open the vowel in order to succussfully escape auditory integration with Z1.

It might be tempting to explain the lowness of F3 in unfiltered tokens in terms of auditory enhancement, in which case a role for F3 in the perception of rhoticity could still be claimed. An F3 that is close enough to F2 such that its auditory correlate Z3 integrates with Z2, as Hayward (2000: 167) describes, would be expected to give rise to a Z_2* having greater perceptual salience than Z2 on its own. That is to say, F3 acoustic energy could be used as a boost, adding its power to that of F2. However, the results from the rhoticity judgement tasks carried out in this study do not support such an explanation. As noted above, filtered tokens were very often judged more rhotic than unfiltered ones, and only rarely judged less rhotic; moreover, apart from two judgements out of forty that filtered *fort* and *hurt* were non-rhotic, the only tokens judged non-rhotic were those containing an acoustic F3. This can be explained if some listeners, a minority of the subjects used in this study, are more sensitive to the suppressing effects of F3 even when it is only weakly present. Sensitivity to these suppressing effects may be determined at least in part by whether listeners have rhoticity in their own speech. Results reported in Yaeger-Dror et al (2009) and Heselwood et al (in press) indicate that listeners' own status as rhotic or non-rhotic speakers can influence their rhoticity judgements, as can perhaps also the extent of their exposure to rhotic speech. In the present study these variables were not controlled for due to the difficulty of ascertaining exposure to rhotic speech. This is a line to be followed in future research.

The low amplitude of F3 seen in the lefthand panel of figure 15, and noted as a general feature of constrictive [ɹ] by Stevens (1998: 542), may well result from an attempt to filter F3 out and thereby to leave Z2 undisturbed, rather than an attempt to enhance it. Because F3 has a low amplitude when its frequency is low in productions of [ɹ], this can only be properly tested using synthetic tokens and varying the frequency and amplitude of F3 independently, but the indications are that when F3 has a high amplitude it does not enhance rhoticity but seems to act to suppress it. If so, then it is the presence of a sharp and distinct Z2 that can be identified as the auditory correlate of rhoticity. Hayward is therefore right in saying that it is a strong peak in the middle of the range of possible acoustic F2 values that gives vocalic rhotics their unique timbre, but wrong in claiming that F3 contributes to the perceptual efficacy of this peak.

Understanding more about how formant relations determine perception of rhoticity, and being able to pinpoint perceptual thresholds and their acoustic correlates that separate r-like perceptions from schwa-like perceptions, and from perception of r-less monophthongs, should contribute signicantly to our appreciation of diachronic changes such as the loss of non-prevocalic constrictive [r] in modern English and the emergence of its more vocalised reflexes (Heselwood, 2009), and also to our appreciation of synchronic variation between r-ful and r-less forms. Future research needs to combine articulatory investigations using techniques such as articulography and ultrasound (e.g. Scobbie, 2009) with acoustic and psychoacoustic methods to explore how they correlate in different phonological contexts with a view to establishing, as precisely as possible, rhoticity thresholds in articulatory, acoustic and auditory domains.

References

Bladon, A. (1983) Two-formant models of vowel perception: shortcomings and enhancements. *Speech Communication* 2, 305-313.

Chistovich, L.A. & Lublinskaja, V.V. (1979) The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research* 1, 185-195.

Engstrand, O., Frid, J. & Lindblom, B. (2007) A perceptual bridge between coronal and dorsal /r/. In Solé, M-J., Beddor, P.S. & Ohala, M. (eds.) *Experimental approaches to phonology*. Oxford: Oford University Press. Pp.175-191.

Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S. & Alwan, A. (2000) Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America* 108, 343-56.

Fujimura, O. & Erickson, D. (1997) Acoustic phonetics. In Hardcastle, W.J. & Laver, J. (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. Pp.66-115.

Hayward, K. (2000) *Experimental phonetics*. London: Longman.

Heselwood, B. (2009) R-vocalisation, linking R and intrusive R: accounting for final schwa in RP English. *Transactions of the Philological Society* 107, 67-98.

Heselwood, B., Plug, L. & Tickle, A. (in press) Assessing rhoticity using auditory, acoustic and psychoacoustic methods. In Heselwood, B. & Upton, C. (eds.) *Proceedings of methods XIII*. Frankfurt: Peter Lang.

Johnson, K. (2003) *Acoustic and auditory phonetics*. Malden, MA: Blackwell.

Ladefoged, P. (2003) *Phonetic data analysis*. Oxford: Blackwell.

Lindau, M. (1985) The story of /r/. In Fromkin,V.A. (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*. Orlando: Academic Press. Pp.157-168.

Plug, L. & Ogden, R. (2003) A parametric approach to the phonetics of postvocalic /r/ in Dutch. *Phonetica* 60, 159-186.

Rosen, S. & Howell, P. (1981) *Signals and systems for speech and hearing*. London: Academic Press.

Scobbie, J. (2009) Ultrasound-based tongue root imaging and measurement. *International Workshop on Pharyngeals and Pharyngealisation*, University of Newcastle-upon-Tyne, 26-27 March 2009.

Stevens, K. N. (1998) *Acoustic phonetics*. Cambridge Mass.: MIT Press.

Stevens, K.N. & House, A.S. (1961) An acoustical theory of vowel production and some of its implications. *Journal of Speech & Hearing Research* 4, 303-320.

Traunmüller, H. (1990) Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97-100.

Wells, J.C. (1982) *Accents of English*. Cambridge: Cambridge University Press.

Yaeger-Dror, M., Kendall, T., Foulkes, P., Watt, D., Harrison, P., Johnson, D., Kavanagh, C. & Oddie, J. (2009) Perception of 'r' by trained listeners. Paper presented at the *Linguistics Society of America Annual Meeting*, San Francisco, 8-11 January 2009.

Barry Heselwood
Dept of Linguistics and Phonetics
University of Leeds
Leeds
UK

b.c.heselwood@leeds.ac.uk

Appendix

	Hz and Bark values at measuring points on spectrograms			
FORT	1	2	3	4
F1	493	516	499	476
F2	819	945	1020	1249
F3	2258	2109	1885	1765
Z1	4.86	5.06	4.91	4.71
Z2	7.37	8.19	8.65	9.90
Z3	13.82	13.37	12.61	12.17
Z1-Z2	2.51	3.13	3.74	5.19
Z2-Z3	6.45	5.18	3.96	2.27
STARS				
F1	562	573	493	447
F2	1117	1278	1444	1547
F3	2321	2206	2110	1868
Z1	5.44	5.53	4.86	4.45
Z2	9.20	10.05	10.84	11.30
Z3	14.01	13.67	13.37	12.55
Z1-Z2	3.76	4.52	5.98	6.85
Z2-Z3	4.81	3.62	2.53	1.25
HURT				
F1	516	510	499	516
F2	1473	1461	1421	1501
F3	2046	1892	1937	1993
Z1	5.06	5.01	4.91	5.06
Z2	10.97	10.92	10.74	11.10
Z3	13.16	12.64	12.80	12.99
Z1-Z2	5.91	5.91	5.83	6.04
Z2-Z3	2.19	1.72	2.06	1.89
NURSE				
F1	430	521	533	447
F2	1645	1622	1616	1662
F3	2430	2332	2338	2372
Z1	4.29	5.10	5.20	4.45
Z2	11.70	11.61	11.59	11.77
Z3	14.31	14.04	14.05	14.15
Z1-Z2	7.41	6.51	6.79	7.32
Z2-Z3	2.61	2.43	2.46	2.38