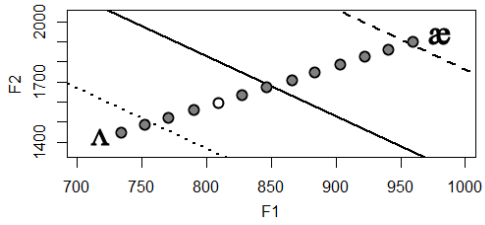
THE ASSOCIATION BETWEEN SPEAKER-DEPENDENT FORMANT SPACE ESTIMATES AND PERCEIVED VOWEL QUALITY

[1]Santiago Barreda, and Terrance M. Nearey
Dept. of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Canada, T6G 2E7 [1]sbarreda@ualberta.ca

1. INTRODUCTION

The term *normalization* will be used here to denote the mechanism listeners use to accommodate between-speaker variation so that they can better identify vowels. Many theories of normalization suggest that vowels are interpreted relative to a representation of a speaker-dependent formant space, rather than being interpreted in an absolute way (Ladefoged & Broadbent 1957, Nearey 1989). Using a representation of a speaker's formant space, vowel sounds may be compared to the expected pattern for each vowel category for that speaker. There is good evidence that the formant spaces of speakers of the same language vary primarily by a single multiplicative parameter (Nearey 1978, Turner et al. 2009). This parameter is closely related to speaker vocal tract length and will be referred to as FF-scaling. Differences in speaker-dependent formant space estimates will be discussed in terms of differing FF-scaling estimates. Speakers with a relatively higher FF-scaling produce higher formant frequencies (FFs) overall than speakers with lower FF-scalings. As a result, when a speaker has a relatively higher FF-scaling, the expected FFs for all vowel categories increase, as do the FFs associated with the boundaries between any given set of vowel categories.

Fig. 1. Points indicate steps along a vowel continuum ranging from a less open vowel (/ʌ/) to a more open vowel (/æ/). Diagonal lines indicate boundaries between vowel categories as determined by varying FF-scaling estimates.



In experiments that involve vowel continua, different apparent speaker FF-scalings can result in vowel category shifts. Consider the simplified situation (ignoring F3 and higher formants) depicted in Figure 1. The dotted line represents the boundary between /ʌ/ and /æ/ when the speaker has a low FF-scaling. If this is the case, then the majority of vowels on the continuum will be identified as instances of /æ/, since they fall above the boundary. On the other hand, if the speaker has a high FF-scaling (as indicated by the dashed line), the majority of vowels along the continuum will be heard as instances of /ʌ/.

If normalization is driven by a process in which listeners estimate the location of the speaker's formant space (based on the apparent FF-scaling of the speaker), the interpretation of a sound with a given set of FFs (represented by a fixed point in Figure 1) should be determined by the FF-scaling estimate arrived at by the listener. Furthermore, the shift in vowel quality should be predictable based on the FF-scaling estimate. In the example given in Figure 1, the white point on the continuum would be more likely to be identified as an /æ/ when FF-scaling estimates were relatively low, indicating an inverse relationship between vowel openness and FF-scaling estimates.

In the experiment to be outlined here, listeners were first trained to report apparent speaker FF-scaling using the training method outlined in Barreda & Nearey (2011). This training method uses voices which vary in f0 and average FFs to teach listeners to report the psychoacoustic quality associated with higher overall FFs. Since higher FFs are associated with a higher FF-scaling, this response variable should be correlated with the apparent FF-scaling of the speaker and, consequently, with the speaker-dependent formant-space estimate arrived at by the speaker. After training, listeners performed a perceptual task similar to that in Barreda & Nearey (2012), in which they were presented with isolated vowel stimuli and were asked to indicate, on each trial: 1) The category of the vowel, 2) The gender of the apparent speaker and 3) Their FF-scaling estimate.

2. METHOD

Participants were 25 native speakers of Canadian English from the University of Alberta. Participants were drawn from a participant pool in which undergraduate linguistics students take part in experiments in exchange for partial course credit.

During the training phase, listeners learned to report apparent FF-scaling using the training method outlined in Barreda & Nearey (2011). After training, listeners proceeded to a testing phase. During the testing phase, listeners were presented with fully-randomized, isolated-vowel stimuli. For each vowel, listeners were asked to indicate the vowel category the stimulus belonged to (either /ʌ/ or /æ/) and the apparent gender of the speaker. Listeners were also asked to indicate the FF-scaling of the speaker using a discrete, 5-point scale, with higher values indicating a higher FF-scaling. Listeners heard each unique vowel stimulus 6 times, resulting in a maximum of 270 responses collected from each listener.

The testing stimuli consisted of a five-step F1-F2 continuum which spanned from FFs roughly appropriate for the /ʌ/ of an adult male to the /æ/ of an adult female. The vowels of the continuum varied in terms of increasing openness, with increasing F1 and F2 frequencies generally resulting in the perception of a more open vowel. The third point in the continuum had FFs which were appropriate for an /æ/ when produced by an adult male or an /ʌ/ when produced by an adult female. Since F1 and F2 frequencies are perfectly correlated, this factor will simply be referred to as F1. The FF values of each step along the continuum is provided in Table I. Each point along the continuum was combined with 3 f0 values (140 Hz, 198 Hz, and 280 Hz), and three F3 values (2475 Hz, 2774 Hz, 3109 Hz), resulting in 45 unique vowel stimuli. All vowels had f0s which decreased linearly by 10% from the start to the end of the vowel, and were 200 ms in duration.

Table 1. Formant frequencies of stimulus vowels.

Step #	1	2	3	4	5
F1	718	775	838	905	977
F2	1422	1536	1659	1792	1935

3. RESULTS

To confirm that listeners were reporting apparent FF-scaling in a consistent manner based on the stimulus properties, a linear model was fit to the pooled data across all participants in which reported FF-scaling was the dependent variable. Stimulus F1, F3 and f0 were the independent variables, and all were coded as continuous covariates. This model explained 18% of the variance in reported FF-scaling, with F1 accounting for 67.8%, f0 accounting for 28.1%, and F3 accounting for only 0.2% of the explained variance.

Table II. Results of significance tests carried out on the within-participant logistic regression coefficients.

Coefficient	Mean	t(24)	p
F1	3.45	16.6	< 0.001
F3	-1.83	11.3	< 0.001
f0	-0.74	5.9	< 0.001
Maleness	0.47	2.6	0.015
FF-scaling	-0.20	2.9	0.009

To investigate the relationship between vowel openness and apparent speaker FF-scaling, a two-stage (Lorch & Myers 1990) logistic regression analysis was carried out. A logistic regression model was fit to the data collected from each participant individually. In each case, vowel openness was the dependent variable, where responses of /æ/ were coded as 1 and responses of /ʌ/ were coded as 0. The stimulus properties F1 step, and F3 and f0 level were coded as continuous covariates. The response variable reported speaker gender was coded as a dummy variable, while reported speaker FF-scaling was coded as a continuous covariate. A series of independent-sample t-tests were carried out on the coefficients collected from all participants to see which independent variables significantly affect perceived vowel openness. The results of this are presented in Table II.

4. DISCUSSION

As seen in Table II, reported FF-scaling has a significant negative effect on vowel openness. This means that for a given vowel sound, when listeners reported a higher FF-scaling, they were less likely to hear an open vowel. This negative association exists despite the fact that vowel openness and reported FF-scaling have a positive marginal relationship. For example, listeners heard /æ/ in 42% of cases when they reported the lowest FF-scaling level, and in 76% of cases when they reported the highest FF-scaling level. As discussed in the introduction, this counter-intuitive result is what would be expected if listeners were normalizing vowels based on speaker-dependent formant space estimates (driven by FF-scaling estimates). Although, in general, vowels with higher FFs will be perceived as indicating a higher FF-scaling *and* a more open vowel, after controlling for stimulus FFs (i.e., considering a fixed continuum point in Figure 1), a higher FF-scaling estimate results in the perception of fewer open vowels overall.

If the association between vowel openness and reported FF-scaling were simply a result of formant estimation errors on the part of the listener, we would expect a positive relation between vowel openness and reported FF-scaling. For example, for a vowel with a given set of FFs, relative to a fixed boundary, in cases where listeners overestimated the FFs, they would be more likely to hear a more open vowel *and* they would be more likely to report a higher FF-scaling.

REFERENCES

Barreda, S. & Nearey, T. (2011). Training listeners to report fundamental frequency and formant range information independently. The 161th meeting of the Acoustical Society of America, Seattle, WA.

Barreda, S. & T.M. Nearey. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *Journal of the Acoustical Society of America* 131: 466-477.

Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels,". J. Acoust. Soc. Am. 29, 98–104.

Lorch, R. F., and Myers, J. L. (1990). "Regression analyses of repeated measures data in cognitive research," J. Exp. Psychol. Learn. Mem. Cogn. 16, 149–157.

Nearey, T. M. (1978). Phonetic Feature Systems for Vowels. PhD thesis, Indiana University Linguistics Club.

Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. 85, 2088–2113.

Turner, R. E., Al-Hames, M. A., Smith, D. R. R., Kawahara, H., Irino, T., and Patterson, R. D. (2006). Vowel normalisation: Time-domain processing of the internal dynamics of speech. in Dynamics of Speech Production and Perception, edited by P. Divenyi. Amsterdam: IOS Press. pp. 153-170.