EVIDENCE FOR THE PERCEPTUAL RELEVANCE OF VOWEL-INHERENT SPECTRAL CHANGE FOR FRONT VOWELS IN CANADIAN ENGLISH

Terrance M. Nearey
University of Alberta, Edmonton, Canada

ABSTRACT

This study assesses the importance of F1, F2 and vowel-inherent spectral change (VISC) in the perception of front vowels in Canadian English using synthetic stimuli with linear F1-F2 trajectories. Results support earlier findings from our laboratories suggesting that VISC plays an important role in the phonetic specification of these vowels.

BACKGROUND

Previous studies in our laboratories [1, 2] confirmed that formant movement plays a role in the specification of phonetic diphthongs such as /e/, which has a diverging F1-F2 pattern (falling F1, rising F2). Results further indicated that the front lax vowels /ɪ/ and /ɛ/ show a converging F1-F2 pattern (rising F1, falling F2), consistent with the targets suggested by Klatt [3] for synthesis of American English vowels. The present study extends our previous findings using synthetic isolated vowels.

PROCEDURE

A four-factor continuum was constructed spanning the front vowels / i, ɪ, e, ɛ, æ/. The variable stimulus factors were onset F1, onset F2, offset F1 and offset F2 frequencies. Onset F1 ranged from 375 to 525 Hz and F2 ranged from 1700 to 2250 Hz (each in four steps). Offset F1 frequencies were manipulated to produce falling (-70 Hz), flat or rising (+70 Hz) linear trajectories for each onset target. F2 trajectories were generated analogously but with 130 Hz rising and falling excursions from the onset values. Vowels were 300 ms in duration with a linearly falling F0 contour (112 to 95 Hz.) Amplitude rose from 35 to 60 dB in the first 60 ms, declined linearly to 58 dB at 150 ms and tailed off to 45 dB at 300 ms. F3 and F4 were fixed at 2500 and 3500 Hz, respectively.

Eleven Canadian English speakers (with some phonetic training) listened to each of the 120 resulting stimuli six times. (Vowels with falling F1 and falling F2 were not synthesized, nor were vowels with the highest level of F1 and F2 onset.) On each trial, listeners identified the category of the vowel and simultaneously assigned a rating (on a scale of 0 to 9) of the "goodness" of each token as a member of the chosen category.

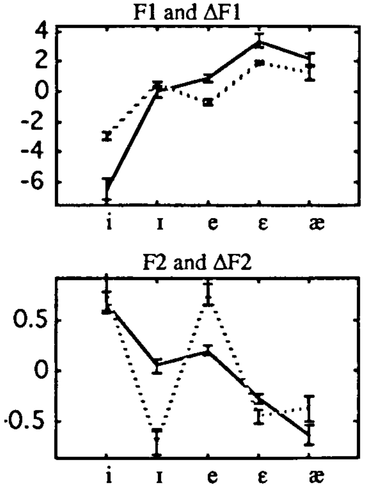

Figure 1. Magnitude (and standard errors) of F1 and F2 coefficients. Solid lines indicate coefficients for onset frequencies, dashed lines for ΔF1, ΔF2.

CATEGORIZATION

Categorization was analyzed using logistic regression, which bears many points of resemblence to the analysis of covariance [4]. Stimulus properties were coded as frequencies (in Hz) of F1, F2, ΔF1 and ΔF2 (corresponding respectively to F1 onset, F2 onset, change in F1 and change in F2 from onset to offset). The coefficients of interest in assessing the relative weight of stimulus properties correspond to vowel-by-stimulus interaction terms. These are plotted as Figure 1. The difference in coefficient values between any two vowels indicates the relative importance of that cue in separating those vowels.

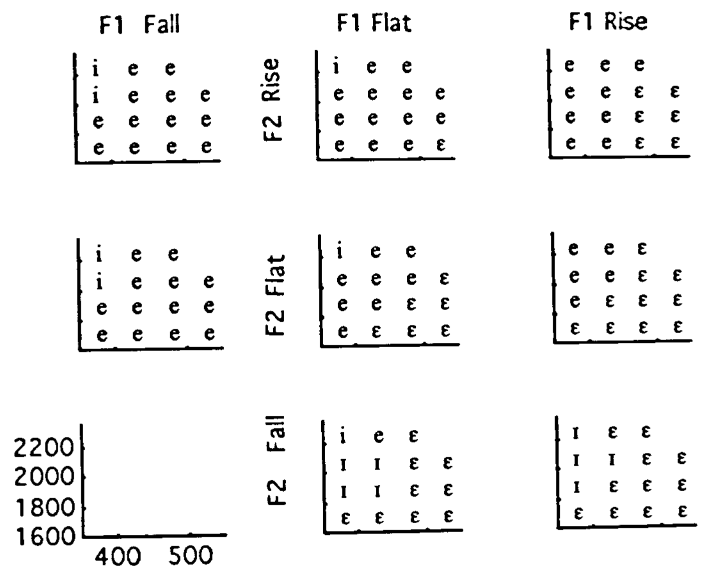

Figure 2. Territorial map for vowels. Columns (left to right): Falling, flat and rising F1 trajectories. Rows (bottom to top): falling, flat and rising F2 trajectories. Stimuli in the lower left corner would possess [w]-like offglides and were not synthesized. Within each cell, the horizontal axis represents F1 (Hz) and the vertical F2 (Hz) of the initial targets of the vowels with frequency values as indicted on the axes of the empty cell.

Random-coefficients regression tests of [5] were applied, testing for the significance of variation in coefficients across vowels compared to inter-speaker differences. Results showed highly significant effects ($p < .0005$) for all stimulus-by-vowel interactions [VxF1: $F(4,7) = 276.2$; VxF2: $F(4,7) = 39.44$; VxΔF1: $F(4,7) = 44.87$; VxΔF2: $F(4,7) = 22.17$.] These tests indicate for F1 and F2 both onset frequencies and changes in frequency are important determinants of listeners' choice of category.

Figure 2 is "territorial map" showing the most frequent response to each stimulus. As expected, formant patterns with falling F1 and rising F2 favored tense /i, e/ responses. There is also a preference for lax /ɪ/ and /ɛ/ with ris-ing F1 and falling patterns. Compatible trends can be observed in the logistic coefficients shown in Figure 1. Vowel categories with relatively higher (positive) coefficient values are favored by a high value of the stimulus property in question over those with lower values. The coefficients of Figure 1 are redisplayed in a 2-dimensional "comet" plot in Figure 3. In this plot (and in Figures 4 and 5) onset frequencies properties are marked with "*" (head of the comet) and offsets by lines (tail). In Figure 3, average offset coefficients have been calculated by adding the ΔF coefficients to the onset coefficients.

The general pattern of the logistic coefficients can be compared to plots like those used in [1] to represent vowel inherent spectral change. Figure 4 shows a comet plot of data from Table II of [1] which is based on the means of production of 10 Canadian English speakers (5 male and 5 female.) Although there are exceptions (discussed further below), the general pattern of means of the onset frequencies is similar as is the direction of movement of formants.

GOODNESS RATINGS

As noted above, "goodness ratings" were also collected for each response. Of primary interest are the locations of patterns where the best tokens of the vowels occurred. A simple estimate of this can be obtained by computing a "center of gravity" of total goodness votes for each vowels in the F1-F2 stimulus space.

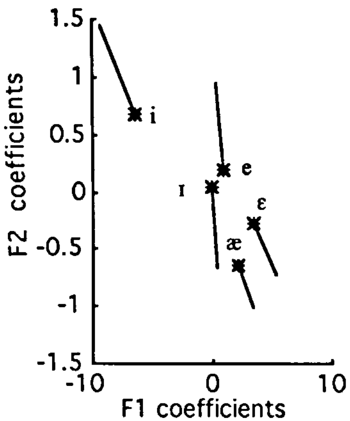

*Figure 3. Plot of mean coefficients from Figure 1 after conversion of ΔF1 and ΔF2 coefficients to frequency offset coefficients. Onset coefficients are marked with *, offsets are unmarked.*

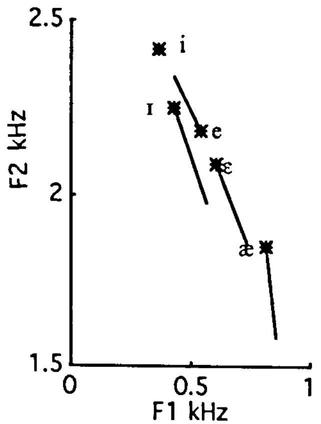

Figure 4 Plot of mean F1 F2 onsets and offsets from Table II of [1]. Labeled ends of lines are onset frequencies and unlabeled tails are offsets.

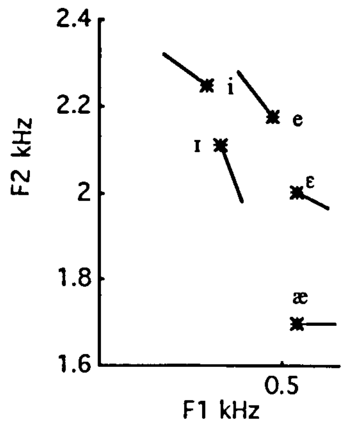

Figure 5 Plot of center of gravity of "goodness votes" for each vowel.

For each of the vowels, the following score was computed:

$$m = \Sigma_i\, x(i)g(i)/\, \Sigma_i g(i),$$

where m is the moment score (in Hz), $x(i)$ is the frequency of the stimulus measure, and $g(i)$ is the total goodness vote for the vowel in question for stimulus i. This total goodness vote is the sum of the total ratings (0-9 scale) across listeners and repetitions. (In order to focus on relatively good stimuli, for each vowel the summation was limited to those stimuli that received a total goodness rating at least 80% of the best rated stimulus for that vowel.)

DISCUSSION

One noticeable discrepancy is that the F1 onset logistic coefficient in Figure 3 for /æ/ is actually somewhat lower than that of /ɛ/, contrary to the trends of the production data in Figure 4. In addition, the vowel /i/ shows fairly large movement toward lower F1 and higher F1 offsets in the logistic coefficients than in the production means.

It should be noted that the onset F1 formant range explored in the synthetic study generally does not extend to those found for /i/ and /æ/ for males. However, these two vowels accounted for only about 7% and 6% of the total responses pooled over subjects and stimuli. Furthermore /æ/ does not show up as the modal response for -- and in fact does not ever receive more than 30% minority response to-- any of the stimuli (see Figure 2). It would seem unwise, therefore to put too much stock in the results for /æ/ without more data in a more appropriate frequency range.

Although /i/ does not account for many responses overall (7%), it does show up as the modal response for 7 of 120 stimuli (see Figure 2). It receives as high as an 85% response for one stimulus (that with lowest falling F1 and highest rising F2) and over 50% for four others. The /i/-pattern for Figure 3 (and 5) is qualitatively compatible with [ij]-like diphthongization often taken to characterize many dialects of North American English (including Canadian dialects [6]) by phonologists. Though we have not found any evidence for such patterns in production of citation forms of Western Canadian English [1, 2], it is not surprising that diverging F1-F2 patterns with low onset F1 and high onset F2 are taken as /i/ by our listeners. Nor is it surprising that the best-rated stimulus patterns for this vowel are those with diverging formants, since the nearest steady-state values are not as extreme as those of production values for male [i].

But a similar argument might be framed by a skeptic who accepts the traditional view that the North American tense vowels are phonetically diphthongal [ij] and [ej], but who believes that /ɪ/ and /ɛ/ are prototypically steady-state. This skeptic might reasonably argue that the categorization results are an artifact of the forced-choice experiment. More specifically, stimuli with converging F1-F2 trajectories are given lax vowel labels because they are more similar to steady-state lax-vowel prototypes than they are to diverging-formant (diphthongal) tense-vowel prototypes.

Even the pattern of results for the goodness data of Figure 5 is not immune to related criticism. However, for one of the lax vowels, /ɛ/ (which received 34% of total responses), there is sufficient relevant data to test the relation between goodness judgments and formant change directly. The stimuli selected for this test were those with flat or converging formants which were also identified by a subject as /ɛ/ six times out of six trials. For eight of the eleven subjects there were sufficiently large numbers of stimuli meeting these conditions that regression equations could be estimated for their responses individually. A (random-coefficients) multiple regression analysis shows a highly significant positive relationship [$t(7)= 4.81$ $p<.01$] between ΔF1 and the goodness rating for /ɛ/ controlling for all other factors. ΔF2 was not significant, though the effects in were in the expected direction ($t(7) = -0.6677$, $p>.5$). Thus there is positive evidence at least for the vowel /ɛ/ that a rising F1 pattern is preferred over a steady state, even when attention is limited to vowels that were consistently identified as /ɛ/. Additional experiments are being planned to explore goodness judgments more thoroughly in local regions of the stimulus space appropriate for each vowel.

ACKNOWLEDGMENT

This work was supported by a grant from SSHRC to the author.

REFERENCES

[1] Nearey, T., & Assmann, P. (1986). "Modeling the role of inherent spectral change in vowel identification." *JASA*, vol. 80, pp. 1297-1308.
[2] Andruski, J., & Nearey, T. (1992). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables." *JASA*, vol. 91, pp. 390-410.
[3] Klatt, D. (1980). "Software for a cascade/parallel synthesizer." *JASA*, vol. 80, pp. 971-995.
[4] Nearey, T. M. (1990). "The segment as a unit of speech perception." *J. Phonetics*, vol. 18, pp. 347-373.
[5] Gumpertz, M., & Sastry, G. (1989). "A simple approach to inference in random coefficient models." *American Statistician*, vol. 43, pp. 203-210.
[6] O'Grady, W.,& Dobrovolsky, M. (1992). *Contemporary linguistic analysis*. Toronto: Copp Clark Pitman.