MODELLING VOWEL INHERENT SPECTRAL CHANGE IN SPONTANEOUS SPEECH

Michelle Sims, Benjamin V. Tucker, and Terrance M. Nearey
Dept. of Linguistics, University of Alberta, AB, Canada, T6G 2E7

1. INTRODUCTION

Phonetic research strongly supports the notion that formant trajectories vary more or less continuously over time even for vowels usually classed as monopthongs (Strange et al. 1983, Nearey and Assmann 1986, Hillenbrand and Gayvert 1993). One possible source for such formant movement is vowel inherent spectral change (VISC). Several studies investigated the role of VISC in vowel perception, arguing that listeners' use F1 and F2 contours in vowel identification. However, this dynamic information has not been investigated in spontaneous speech. It is possible that the VISC- relaed dynamic spectral movement seen in carefuly produced vowels will not be evident in conversational speech due to coarticulatory effects of the surrounding consonants (Strange et al. 1983). The present study investigates this issue by looking at the vowel productions in a corpus of spontaneous speech. Several alternate accounts of the nature of VISC in speech production are also briefly discussed.

1.1 Theories of Vowel Inherent Spectral Change

In their work on the perception of VISC, Morrison and Nearey (2007) discuss three predominate approaches of analysing vowel movement: the onset+offset, onset+slope, and onset+direction hypotheses. The onset+offset, or ΔF, approach holds that a vowel's ending F1 and F2 values are important to the perception of vowels. The onset+slope, or $\Delta F/\Delta t$, approach models VISC as a function of time, stating that modelling vowel movement through time enhances the perception of vowels. The onset+direction model, on the other hand, states that it is the overall direction of movement, not necessarily the offset or slope, that provides a perceptual cue to vowel identity. VISC directions are measured in terms of the rise and fall in F1 and F2 (i.e. F1 rising and F2 falling, F1 falling and F2 rising, no movement, etc.).

Research on vowel identification has suggested that the onset+offset hypothesis best captures the perceptual cues listeners use (Morrison and Nearey 2007, Hillenbrand et al. 2001, to name a few). Studies on the production of vowels in careful citation speech have added to this theory with models that include vowel duration and pitch (Hillenbrand et al. 2001). The present paper tests these approaches to modelling VISC in spontaneous speech. As much of the research on VISC has focused on perceptual studies or citation speech, it is of interest to see if these theories also hold for vowels produced in everyday conversations.

2. METHOD

The present study makes use of a dataset of 54 monosyllabic irregular English verbs that differ between their past and present tense forms based on a single vowel alteration. For example, we included irregular verbs like sing/sang in our dataset, but excluded irregular verbs such as weep/wept and is/were. We extracted the vowels of these verbs from the Buckeye Corpus of Conversational English (Pitt et al., 2007), yielding 7,034 tokens of eleven different monophthongs from 40 adult speakers (20m/20f)

F1, F2, and F3 contours for each vowel were gathered using FormantMeasurer (Morrison and Nearey 2011) and hand-corrected. In accordance with Nearey and Assmann (1986), we marked the onset of each vowel at 24% of its entire duration and the offset at 64%. In doing so, we attempt to mitigate, to some extent, the influence of coarticulation from the surrounding consonantal context.

We performed a cross validation discriminant analysis to test each VISC hypothesis' ability to capture the spontaneous speech data. The analysis is based on a linear parametric technique trained on all various combinations of F1 and F2 onsets, offsets, slope ($\Delta F/\Delta t$), direction (all 9 combinations of F1 and F2 falling, rising and no movement), pitch and duration. We performed the discriminant analysis both on males and females (separately and combined). We then used t-tests to test the significance of the VISC movement and differences across males and females.

3. RESULTS

Figures 1 and 2 show VISC movement for males and females, respectively. On the plots, the arrow indicates the average vowel offset and the labelled end represents the average vowel onset. The linear VISC movement shown is gathered from the onset+offset model. Though the vowel spaces are significantly different between genders ($p<0.001$ or all vowels), the trends in the VISC movements for males are not significantly different from females ($p>0.05$ for all vowels). The slopes for each VISC movement across males and females, are significant ($p<0.005$ for all vowels). It is interesting to note the extreme fronting of /u/ that is characteristic of the Ohioan dialect.

Table 1 illustrates the outcomes of the discriminant analysis. For simplicity, we have only included those models that perform the best. The percentages indicate the amount of improvement each model contributes to vowel

discrimination compared to a model consisting of a single F1 and F2 measurement. These improvements are seen in the separate models for males and females, as well in the combined gender model. Regardless of gender, a model consisting of formant onsets, offsets, pitch, and vowel duration performed the best at the discrimination task.

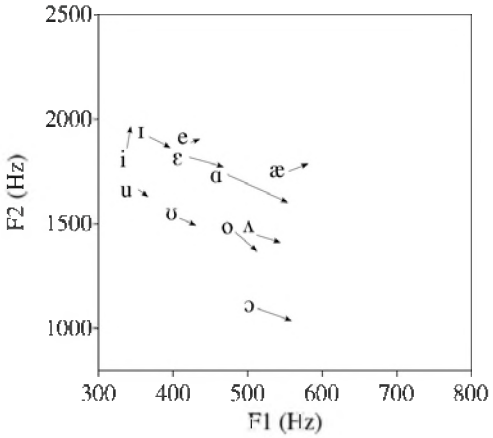

Figure. 1. Average vowel inherent spectral change for males.

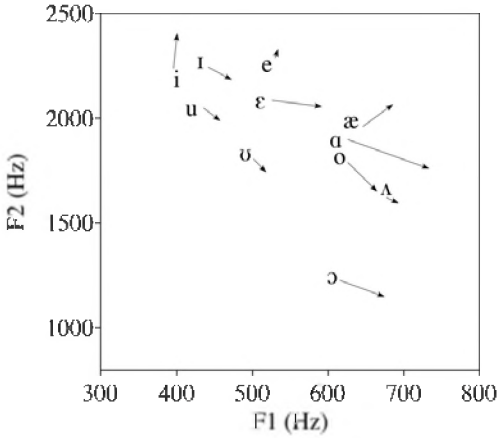

Figure. 2. Average vowel inherent spectral change for females.

Table 1. Results of the discriminant analysis: amount of improvement with the individual addition of Slope, Direction and Offet compared to a Onset+Pitch+Duration model.

	...+Slope	...+Direction	...+Offset
combined genders	6.57%	6.74%	12.35%
males only	4.80%	4.69%	10.30%
females only	7.80%	7.95%	15.77%

4. DISCUSSION

Contrary to theories of dynamic vowel specification, the present study finds evidence for vowel inherent spectral change in spontaneous speech productions. The trends in F1 and F2 frequency shifts in English monophthongs seen here compliments similar findings from carefully produced vowels in a limited consonantal environment. That is to say, we find support for VISC movement in spontaneous speech across a wide variety of phonetic contexts in line with previous research in more controlled conditions.

All in all, our models of VISC in spontaneous speech support the perception research and studies on citation speech. We find a slight superiority for a combined onset+offset+duration+pitch model in capturing the dynamic spectral properties of vowels in conversational English. These results are in line with both Hillenbrand et al.'s (2001) research on citation vowels and other studies comparing the different approaches to VISC analysis (Morrison and Nearey 2007). It is notable that even though the vowel spaces differ between males and females, the trends in VISC movement remain the same. This is evident in both the discriminant and difference tests. Moreover, we find support for the direction and types of formant movement first found by Nearey and Assmann (1986).

These findings have immediate implications for perceptual experimentation. Most studies on vowel identification make use of a onset+offset theory of VISC, with the data here point to the discriminative importance of this hypothesis in spontaneous speech productions. Theories of speech processing, too, could benefit from this more complex and integrated model of dynamic vowel specification.

REFERENCES

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. Journal of the Acoustical Society of America, 109, 748–763.

Hillenbrand, J. M., & Gayvert, R. T. (1993). Vowel Classification Based on Fundamental Frequency and Formant Frequencies. Journal of the Acoustical Society of America, 36, 694-700.

Morrison, G. S., & Nearey, T. M. (2007). Testing theories of vowel inherent spectral change. Journal of the Acoustical Society of America, 122, EL15–EL22.

Morrison, G. S., & Nearey, T. M. (2011) FormantMeasurer: Software for efficient human-supervised measurement of formant trajectories. [Software release 2011-05-26].

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. Journal of the Acoustical Society of America, 80, 1297–1308.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler–Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www. buckeyecorpus. osu. edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America, 74, 695–705.

ACKNOWLEDGEMENTS

Work supported by Social Sciences and Humanities Research Council of Canada Grant number 410-2011-0386 to B. V. Tucker.