Best Practices in Automatic Vowel Production Analysis

Rebecca Roeder (*rroeder@uncc.edu*) & Sky Onosson (*sky@onosson.com*)
University of North Carolina at Charlotte **University of Victoria** @onosson

1 Introduction

This paper 1) provides observations on using FAVE to measure vowel formants with respect to systemic change and 2) suggests an approach to accurate analysis of yod that incorporates FAVE.

FAVE: The University of Pennsylvania Forced Alignment & Vowel Extraction automated software program (Evanini 2009, Rosenfelder et al 2011).

Acoustic features analyzed wrt FAVE:

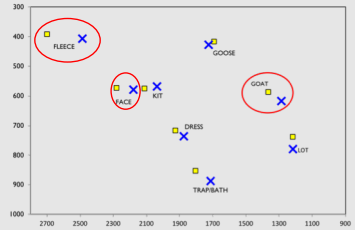
- Measurement point for cross-study comparison
  - SCVE: Synchronic Corpus of Victoria English (D'Arcy 2018)
  - PCE: Phonetics of Canadian English (Boberg 2008, 2010)
- Bandwidth: Monophthongs (SCVE)
- Accurate analysis of yod: Mixed methods (SCVE)
  - Perceptual analysis of yod vs non-yod
  - Acoustic diagnostic for yod-dropping
  - Qualitative vs quantitative observations

2 Measurement point & FAVE

FAVE facilitates only minimal adjustment of measurement point.

Apparent time comparison of vowel production across studies is possible only if measurement pt is same for a given vowel.

Example: Boberg (2008) wordlist, measurement point differs: Victoria, BC (SCVE) vs. Pan-Canadian averages (PCE)



□ = Victoria, age **14-23** (N = 16) X = pan-Canadian average, age **14-23** (N = 84)

FAVE default method used to measure Victoria tokens.

Vowels not circled were measured at the same point in the duration of the vowel in both studies, enabling direct comparison.

Vowels circled in **red** were measured at different points in duration, making direct comparison impossible.

FAVE options: fourth, third, mid, lennig, anae, maxint

"The default method, faav, modifies the **third** method in that /ay, ey/ are measured at maximum F1, /ow, aw/ halfway between maximum F1 and the beginning of the vowel, and /Tuw/ (/uw/ after coronal consonants) at the beginning of the vowel". (Fruehwald 2013)

3 Bandwidth & FAVE

Bandwidth is essential to accurate formant readings, especially for data recorded in non-optimal environments.

Measuring formant frequency: Broadband spectrogram
Formant frequency: "Of a complex sound, a range of frequencies in which there is an absolute or relative maximum in the sound spectrum...The frequency at the maximum is the formant frequency," (ASA 1994).

A formant frequency measurement is by definition a composite measure of the frequency of one spectral peak in the complex wave form resulting from the resonance of the vocal tract.

Broadband spectrograms enable a frequency analysis that is coarse enough to collect energy across the group of frequency components that create the spectral peaks that correspond to formants. Broadband blurs together a large enough band of frequencies to display the collection of frequency components that comprise the broad spectrum envelope peaks that correspond to vocal tract formants.

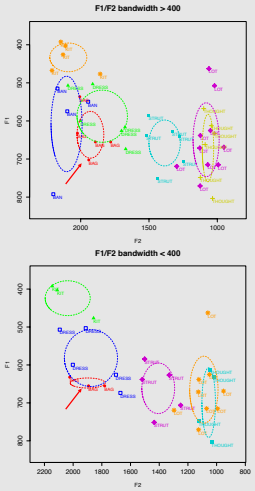
Harmonics (multiples of F0 in the spectrum) also play a role in speech formant production. Formants are strongest where they intersect with harmonics.

Bandwidth: High bandwidth = lack of precision
"the difference between the upper and lower frequencies in a contiguous set of frequencies," (Zutshi 2010: 8).

Bandwidth can increase due to internal factors that affect resonance (e.g., nasalization) or external factors that affect signal-to-noise ration (e.g., background noise).

Monophthongs: FAVE provides F1, F2, F3 bandwidth for single-point measurement.

Bandwidth accuracy example: BJ29m (SCVE)



More tokens, but less accuracy

All 7 vowels shown display modification in average and s.d. due to bandwidth

Greater accuracy, but fewer tokens

Pre-nasal /æ/ difficult to measure accurately due to interference of nasal resonance

Diphthongs: FAVE does not provide bandwidth for interval analysis. However, DARLA does.

4 Analysing Yod-dropping

Not all features are best analyzed via the results obtainable from FAVE, but with some compromise FAVE can still provide useful information.

Identification of yod – no preexisting metric
Yod refers variability in pronunciation of the on-glide /j/. In Canadian English a subset of coronal-initial GOOSE e.g. *new, tune, student*, etc. (hereafter NEW) exhibits variable occurrence of yod, e.g. [nu] vs. [nju].

Studies which have examined yod variation in Canada almost invariably use self-reporting (e.g. Pringle 1985; Chambers 1998) or auditory analysis (e.g. Clarke 1993; Woods 1999) rather than replicable acoustic metrics.

Glide+vowel sequences problematic for segmentation (manual or automatic)

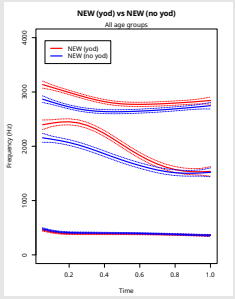
Yod, among glides, is particularly vowel-like (Gick 2003; Padgett 2008). Because of this and the lack of existing metrics, we (Roeder et al 2018) required a novel methodology for its identification in Canadian NEW.

To classify tokens for yod occurrence, a random subset of SCVE NEW tokens were auditioned by three trained sociophoneticians and classified for yod presence or absence. Inter-rater agreement was high at 81.25%, considered reliable (Clopper 2011).

Acoustic analysis of *yod+V* sequences from these tokens revealed that the greatest point of difference between yod retention vs. loss was the value of F2 at 25% of duration. SSANOVA comparisons confirm this finding.

SSANOVA comparison of NEW (yod) vs. NEW (no yod) formants – SCVE data

Greatest significant difference between formants occurs in F2 at 25% duration



FAVE measures formant values at 20%, a reasonable approximation. All potential *yod+V* sequences were tagged as a single segment, and so treated as a unitary vowel by FAVE – essentially a hack, but it works!

FAVE output allowed us to determine yod retention rates in the SCVE, revealing that Victoria retains yod in NEW at a high rate of 39.5% which is moreover stable across gender and age, despite increased yod loss generally throughout Canada (in the U.S. it is essentially complete).

The available output of FAVE (or other automated tool) needs to be considered in design phase

Points to consider in study design:

- FAVE provides formant output at 5 discrete points: 20%, 35%, 50%, 65% and 80% of duration
- For certain types of highly sonorous, vowel-adjacent segments such as glides and liquids, the characteristics best distinguishing consonant from vowel may occur at other positions
- FAVE's output may or may not always be greatly amenable to the particulars of a given research question
- Failure to account for this in planning stages may lead to unsatisfactory results, wasted time, etc.

5 Conclusions

- FAVE allows for only limited manual control over measurement point, posing a potential obstacle to comparison with data sets not measured using FAVE.
- Given that low bandwidth is essential to accurate acoustic measurement of vowel quality, FAVE provides necessary formant information for analysis of monophthongs (i.e., F1, F2, F3 bandwidth at 1/3 of duration) but not diphthongs (no bandwidth readings for interval analysis).
- Yod can be handled reasonably well under FAVE output, provided that the glide+vowel sequence is not further segmented, but other similar phonetic features may not fare so well, depending on their qualities. Researchers are advised to carefully consider the right tool for the job, and how it can be modified or adapted to special-use cases.

References

- Boberg, C. 2008. Regional phonetic differentiation in Standard Canadian English. *Journal of English Linguistics* 36, 2: 129–154.
- Boberg, C. 2010. *The English language in Canada: Status, history and comparative analysis.* New York: Cambridge University Press.
- Boersma, P. & D. Weenink. 2015. Praat: Doing phonetics by computer [Computer program]. Version 5.4.12, retrieved July 2015 from http://www.praat.org/
- Chambers, J. K. 1998. Inferring Dialect from a Postal Questionnaire. *Journal of English Linguistics*, 26, 3: 222–246.
- Clarke, S. 1993. The Americanization of Canadian pronunciation: A survey of palatal glide usage. In S. Clarke (ed.), *Focus on Canada.* Amsterdam: John Benjamins. 85–108.
- Clopper, C. G. 2011. Checking for reliability. In M. Di Paolo & M. Yaeger-Dror (Eds.) *Sociophonetics: A student's guide.* New York: Routledge. 188–197.
- D'Arcy, A. 2018. *Linguistic variation and language change: Synchrony meets diachrony.* Cambridge: Cambridge University Press.
- Fant, G. 1960. *Acoustic theory of speech production.* Mouton & Co.: The Hague, Netherlands.
- Fruehwald, J. 2013. FAVE-extract. Last accessed June 2018 from https://github.com/JoFrhwld/FAVE/tree/master/FAVE-extract
- Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic Interpretation: Papers in Laboratory Phonology VI*, 222–236.
- Nearey, T. M. 1978. *Phonetic feature systems for vowels.* Bloomington: Indiana University Linguistics Club.
- Padgett, J. 2008. Glides, vowels, and features. *Lingua*, 118, 12: 1937–1955.
- Pringle, I. 1985. Attitudes to Canadian English. In S. Greenbaum (Ed.), *The English language today: Public attitudes to English* (pp. 183–205). Oxford: Pergamon.
- Roeder, R., Onosson, S. & A. D'Arcy. 2018. Joining the western region: Sociophonetic shift in Victoria. *Journal of English Linguistics* 46, 2: 87–112.
- Rosenfelder, I., J. Fruehwald, K. Evanini & J. Yuan. 2011. FAVE (Forced Al lignment and Vowel Extraction) program suite. Retrieved Oct 2014 from http://fave.ling.upenn.edu
- Woods, H. B. 1999. *The Ottawa survey of Canadian English.* Kingston, ON: Strathy Language Unit, Queen's University.

* The data on which this study is based were extracted from the Victoria English Project funded by the Social Sciences and Humanities Research Council of Canada grant no. 410-2011-0219 to Alexandra D'Arcy.