Assessing automated methods of analyzing vowels in sociophonetics

Lanlan Li
University of Manitoba

Two computational tools have been developed to help sociophoneticians to obtain vowel alignments and extract formant values automatically: one is semi-automated FAVE (Rosenfelder et al., 2011), another is completely automated DARLA (Reddy and Stanford 2015) which doesn't require human transcriptions. The main purpose of this project is to examine and present their performance on detecting vowel variations in an English ethnolect spoken by a Filipino community in Winnipeg, Canada. At the same time, comparisons regarding workflows, average time spent, input and output systems between the two automated methods were conducted. Previous research analyzing wordlist data shows a non-significant generational difference in terms of Canadian Shift vowels (hypothesis 1), but a significant difference in F2 between older and younger generation for /u/ ($p=<.01$), meaning younger speakers are fronting /u/ more than older speakers (hypothesis 2). In this case, the two hypotheses were used to test performance of FAVE and DARLA.

The data come from 14 sociolinguistic interviews of 7 English L1 younger (≤40, 5F, 2M) and 7 English L2 older (≥40, 5F, 2M) speakers who are respectively the second and first generation of Filipino immigrants in Winnipeg, Canada. Vowel data from FAVE were aligned and extracted by locally-installed FAVE-align and FAVE-extract Toolkits. DARLA data were returned by DARLA online interface after uploading clipped audio without the interviewer's sound. 20 stressed vowels for each phoneme and each speaker were randomly selected. We focused on Canadian Shift vowels /ɪ, ɛ, æ/ (Clarke et al., 1995, Boberg, 2005) and *u* in three phonological environments (following coronals, preceding laterals, and elsewhere, Podesva 2011). Mixed effects modeling in R (R Core Team 2015) were conducted with *F1, F2,* and *agegroup* as fixed effects, and *lexical item* and *speaker* as random effects.

Results for the 13 vowel phonemes show absolute mean differences between the two methods with F1 ranging from 1 Hz to 38 Hz, and from 5 Hz to 185 Hz for F2. Statistical models reveal that *agegroup* is not a predictor of F1 or F2 of Canadian Shift vowels for the data produced by either FAVE or DARLA. This supports and agrees with hypothesis 1. Regarding /u/, FAVE supports hypothesis 2 and presents a statistically significant difference between the two generations ($p = .0005$***), indicating that younger speakers are more fronted than the older ones. This pattern is also seen in DARLA's ($p = .015$**). Moreover, when /u/ data is split according to the three phonological conditions, both systems reveal similar results presenting TOO in the most fronted position.

This study shows that furthermore although overall time outlay is higher with FAVE than DARLA, FAVE is better able to detect sociophonetic vowel variation in our ethnolectal data, exceeding DARLA's performance regarding high back vowel fronting. For DARLA, even though our results show its low accuracy on vowel transcriptions (overall 45%), phonetic analyses of DARLA data are consistent with FAVE's.