An Evaluation of Sociolinguistic Elicitation Methods

Zac Boyd,[i] Zuzana Elliott, Josef Fruehwald, Lauren Hall-Lew, and Daniel Lawrence

The University of Edinburgh

ABSTRACT

This study investigates the effects of different elicitation methods on the speech of a single speaker of San Francisco English who is participating in a systematic set of vocalic sound changes known as the California Vowel Shift [6]. We contrast data obtained from classic sociolinguistic interview methods with data from self-recordings, as well as data from various methods for eliciting spontaneous speech that are typically used in laboratory settings.

An analysis of five sound changes indicates that self-recorded speech often results in significantly more advanced productions than interview speech, while speech from laboratory methods is largely comparable to interview speech. Surprisingly, differences between read speech and unscripted speech are minimal. We conclude by recommending the utility of controlled-but-spontaneous laboratory elicitation methods, and by strongly recommending the use of self-recorded data for studies of sound changes in progress.

Keywords: intraspeaker variation, vowel quality, naturalistic speech, California English, methods

1. INTRODUCTION

Sociolinguistic interviews often aim to capture a range of different speech styles from an single informant. This is typically achieved through the inclusion of an extended conversational section followed by a series of reading tasks such as a reading passage, word list, and a minimal pair list [14]. A common secondary aim of these read tasks is to obtain tokens of variables which occur relatively infrequently in naturalistic speech.

Despite the need for careful choice of tasks in achieving these two aims, there is surprisingly little discussion on how to choose or design a reading task, nor advice on alternative methods for obtaining productions of specific variables. Reading tasks are often not ideal in fieldwork contexts where some informants might be pre-literate, semi-literate, or have poor eyesight. Given such challenges, it is perhaps surprising that the latest textbooks on sociolinguistic research methods (e.g. [13, 15]) do not detail the growing number of ways of obtaining spontaneous speech using controlled elicitation methods. Furthermore, no empirical studies exist validating the efficacy of different sociolinguistic elicitation methods. The present study aimed to address this goal while providing recommendations for task design in future work.

The analysis is based on data obtained in a context akin to a classic sociolinguistic interview, augmented with participant self-recordings (made without the interviewer present) as well as various interviewer-led laboratory methods for eliciting spontaneous speech. We focus on a single speaker who is part of a larger sociophonetic examination of sound changes in the English of San Francisco, California. This set of changes has been termed the California Vowel Shift (CVS) in previous literature [6, 9, 10, 11, 12, 17], and comprises the systematic lowering and backing of the lax front vowels (KIT, DRESS, TRAP) and the fronting of the mid- and high back vowels (GOAT, GOOSE), all of which are considered here. Additionally, TRAP before nasals (BAN) is also raising and fronting. The CVS also involves the merger of the low back vowels (LOT, THOUGHT) and the fronting of FOOT and STRUT, but these lexical sets are not analysed in the present paper due to time and space constraints.

The research questions in the present paper focus on self-recordings on the one hand (e.g. [18, 21]), and laboratory methods on the other (e.g. [1, 3]), and the extent to which each results in significantly different vowel qualities from the baseline, conversational interview speech. We also consider the use of reading passages. Overall, we find that self-recordings result in significantly more advanced variants than either interview speech or laboratory task speech, but that there are few significant differences between interview speech and laboratory task speech. We also find few differences those two styles and read speech; self-recordings stand out as markedly different from all the other speech styles. This finding leads us to two practical recommendations: first, that sociolinguistics and especially sociophoneticians should strongly consider obtaining self-recordings whenever possible, and second, that the growing range of spontaneous speech elicitation tasks born in laboratory settings can be usefully brought into the toolkit of the sociolinguistic fieldworker without introducing significantly different speech styles.

2. METHODOLOGY

Primary data for the study come from recordings of a bilingual Chinese American female from San Francisco, born in 1985 and recorded in 2013. This speaker participated in an extended sociolinguistic interview, which included seven reading passages in order to acquire a greater amount of read speech than is usually obtained in such an interview:

1. The Rainbow Passage [7]
2. The North Wind and the Sun [5]
3. Please Call Stella [24]
4. The Wolf Passage [5]
5-7. Basketball, Dolls, Victoria's Secret [8]

The interview itself was relatively brief (10 min) and no word list or minimal pair list was included. Rather, data elicitation focused on testing the viability of bringing laboratory tasks into a field setting. Following the interview session the speaker participated in these six tasks, which we will refer to as 'lab tasks' since most of them originate from phonetics work conducted in laboratory settings:

1. An interactive Map Task [3]
2. An interactive Diapix task [1]
3. A monologic Diapix task, with an imagined interlocutor; e.g., [20]
4. A picture book narration [23], as in [22]
5. A silent movie narration [4]
6. Semantic Differential questions [14]

All methods have been used in various laboratory settings (and sociolinguistic interview settings, in the case of the Semantic Differential questions) where the focus was on obtaining controlled but naturalistic speech for purposes of phonetic analysis. The one adaptation we made here was (3), bringing a monologic speech style to the typically dialogic Diapix task. This was done in an attempt to tease apart the cause of any potential differences between tasks as being due to the difference between monologue and dialogue versus a difference between individual tasks.

The speaker also carried out three self-recordings during the course of the same week: one prior to the extended interview (a visit with two friends) and two following it (lunch with a friend and cooking at home with her sister). She was recorded in public and private settings, including but not limited to cafes, restaurants, and her home. The interview and the lab tasks were led by the fourth author (also a friend of the speaker), who acted as the interlocutor in the interactive tasks. Recordings were made using a Zoom H2 recorder with an external lapel microphone and were digitized at 44,100 kHz.

All recordings were transcribed orthographically using ELAN [15], resulting in a corpus of 18,004 words. The procedure resulted in 2703 target words with KIT, DRESS, BAN, TRAP (preceding orals only), GOAT, and GOOSE. Table 1 summarizes the number of tokens for each variable by style (interview, self-recording, laboratory-based, and reading passage).

Table 1: Token counts for all variables by style

Vowel	Interview	Self-Rec	Lab	RdgPssg
KIT	42	297	330	51
DRESS	52	291	497	102
BAN	81	105	319	39
TRAP	55	302	337	78
GOAT	68	312	304	48
GOOSE	17	120	196	86

Because of the time and energy constraints on the participant herself, the interview portion of the extended interview was much shorter than in usual sociolinguistic field settings. As a result, the self-recordings and lab tasks elicited substantially more data than the interview. Both the self-recordings and the lab tasks have many times more data than the interview and reading passage styles. The data deficit of the interview is most clear for GOOSE, which only had 17 tokens in the interview; statistical analysis of this variable should be treated with particular caution.

2.1 Acoustic Methods

Automatic alignment and vowel extraction was conducted using FAVE [19]. We analysed F1 and F2 at 20% of the vowels' duration, as provided in the FAVE output. Tokens were excluded where automatic measurements failed for F1 or F2 at 20% (N=17), where tokens were immediately preceded (N=42) or followed (N=8) by laughter, noise, audible breaths or lip smacks, and where tokens were immediately preceded (N=761) or followed (N=713) by a vowel or a pause.

2.2 Descriptive Analysis and Defining Speech Styles

The data from three self-recordings were combined and analyzed together, as were the measurements from the seven reading passages and the six lab tasks
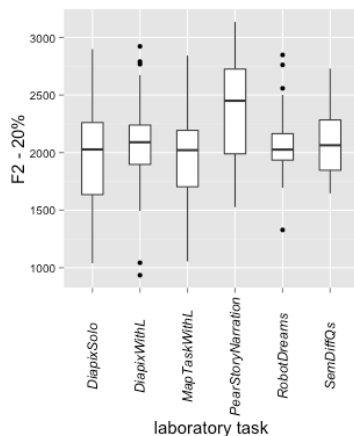
The amount of variation across different reading passages differed by variable. Some variables (KIT F1, DRESS F1) show few differences between reading passages. Other vowels show a very large amount of difference across passages (KIT F1, DRESS F2, GOOSE F1 and F2), while the rest are middling.

Figure 1 gives an example of the variation across different reading passages. While the differences highlight that the nature of a study's 'reading passage style' depends very much on the reading passage itself, it may be that the differences seen here are readily explained by well-known linguistic constraints (e.g., phonological context or lexical frequency). Future analyses will take a closer look at this level of variation, but for the remainder of the analysis these differences are collapsed and all reading passages are treated as one style.

Figure 1: TRAP F2 across 7 reading passages
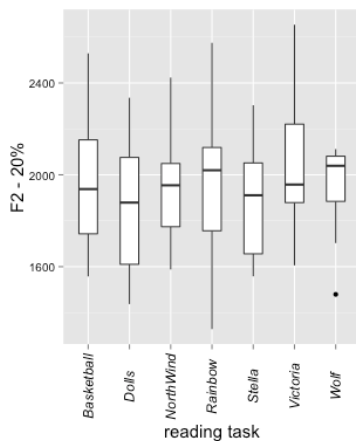
Figure 2: DRESS F2 across laboratory tasks

The amount of variation across different lab tasks also differed by variable. Some show very few differences between tasks (TRAP F1 and F2), even between the spontaneous and read tasks. Other vowels show very large difference across tasks (DRESS F1, GOOSE F2), while the rest are middling (KIT F1, BAN F1 and F2). Figure 2 gives an example of the large amount of variation that might be seen across different lab tasks. Note that 'Pear Story Narration' [4] refers to the silent movie narration task, and 'Robot Dreams' [23] refers to the picture book narration task. Again, for the purposes of the present study, data from all six tasks are combined to represent 'lab speech'.

3. ANALYSIS

Mixed-effects linear regression analyses were conducted for each variable using the *lme4* [2] package in R, with 95% confidence intervals estimated by 1,000 parametric bootstrap replicates using *bootMer*. In addition to the elicitation method and location in the word (initial, final, internal, co-extensive), the preceding and following segments were included as fixed effects for all models. Random intercepts by word were also included.

Predictors included in the initial models were the place and manner of articulation of the preceding and following segments, and speech style (interview, laboratory, reading passage, and self-recording). Subsequent models were fit on a subset of the data comparing only the laboratory tasks. These included the same linguistic constraints as above and a seven-level factor representing the task.

4. RESULTS

With respect to sound change, we expect reading passage speech to function as the most conservative style of these four styles. However, we do not see much evidence in these data for the specific contrast between reading passage speech and interview speech. Rather, the most robust findings obtained are between the lab tasks and interview speech, on the one hand, and to an even greater extent between the self-recordings and interview speech. This is perhaps surprising since many of the lab tasks also contain a read component, whereas the interview and self-recordings are both entirely non-scripted.

Figures 3a and 3b plot the differences obtained between lab speech, self-recordings, and interview speech (baseline; confidence intervals include 0). For the most part, self-recordings and lab elicitation methods did not differ reliably from interview speech on F2. The exceptions are:
- KIT – fronter in lab speech; less advanced
- TRAP – backer in self-recording; more advanced
- GOAT – fronter in both styles; more advanced

However, on F1, all three CVS vowels are reliably lower in self recordings from interview speech (Figure 3b). There is no reliable height difference between lab speech and interview speech for these variables. The self-recordings result in reliably different vowel qualities than the interview speech in the following ways:
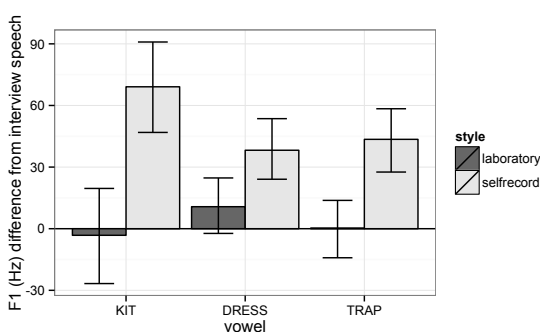- KIT – lower; more advanced
- DRESS – lower; more advanced
- TRAP – backer; more advanced
- GOOSE – backer; less advanced

Self-recording is also significantly different from interview speech for BAN, which is lower and backer in the self-recordings, i.e., less advanced in the CVS.

Figure 3a: F2 variation by task type
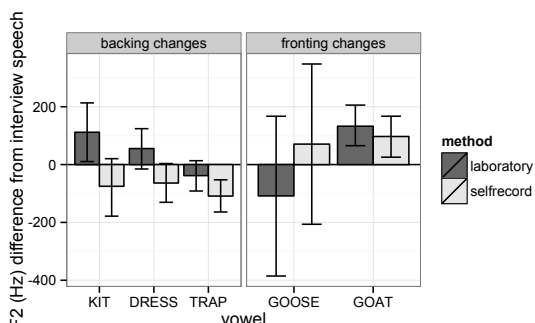
Figure 3b: F1 variation by task type

Overall, we find one area of evidence for those expected effects of stylistic difference: more advanced variants of the California Vowel Shift are usually produced in self-recorded speech than in interview speech. Furthermore, we find encouraging evidence that methods that have originated in laboratory contexts translate very well to field-based interview conditions, without resulting in gross significant differences in phonetic production. However, the data do not show any of the expected differences between read speech and spontaneous speech. In addition, these differences do not apply to all of the vowel quality contrasts under study.

5. DISCUSSION

The two-fold goals of the standard sociolinguistic interview with respect to language change are:
- To collect a sufficiently large volume of speech that will allow for statistical analysis of internal and external factors on linguistic variation, and,
- To collect speech that is as naturalistic as possible while also often attempting to obtain productions of naturally low frequency words.

The question is how well the various tasks examined here address these goals.

Interviews are the standard baseline method, and in comparison to the typically used read speech tasks (reading passages, word lists, minimal pair lists) interviews are considered to elicit the most naturalistic or vernacular speech, which in cases of sound change is taken to mean the style with the most advanced productions for a given speaker. In contrast, reading passages are expected to elicit more conservative productions. In the present paper, neither is true: interview speech does not show the most advanced tokens, and reading passages appear to pattern with speech that is stylistically similar to interviews.

Self-recordings have been increasingly used in sociophonetic work in the recent decade [18, 21]. They present a considerable methodological challenge in that they require extra help on the part of the participant, who also has to be trained in making recordings. However, our results suggest that the payoff is substantial. Compared to an analysis of interview data, the participant studied here appears to be a more advanced speaker of the California Vowel Shift on most of the vowel qualities analyzed. Self-recordings outperform interviews in this regard, although their ecological validity means that the elicitation of particular word forms is impossible.

A happy medium lies in the methods we are calling 'lab tasks', which have proved useful in laboratory studies for decades, but usually do not enter into field-based sociolinguistic methods. The evidence here suggests that they are ideal for achieving both of the main goals in sociophonetic research, especially for studies of sound changes in progress, in that any weaknesses they have seem to be shared nearly equally with classic interviews.

6. CONCLUSION

Expanding the toolkit of elicitation methods has many benefits to sociolinguists. We suggest that the methods profiled here enable the researcher to collect a large volume of naturalistic data in a more controlled and replicable way than classic interview methods allow. As there is little quantitative difference found between the data collected in laboratory tasks and that collected through interview speech, the inclusion of a wide array of controlled elicitation methods does nothing but aid the researcher in providing ample data for analysis with a range of potential stylistic variability. Furthermore, self-recorded speech provides evidence of the most advanced variants of the vowels examined here and should be incorporated in sociophonetic methodology whenever practically possible.

7. REFERENCES

[1] Baker, R., Hazan, V. 2011. DiapixUK: A task for the elicitation of spontaneous speech dialogs. *Behavior Research Methods*, 43(4), 761-770.

[2] Bates D, Maechler M, Bolker B and Walker S (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

[3] Brown, G., Anderson, A. H., Shillcock, R. and Yule, G. 1984. *Teaching Talk: Strategies for production and assessment.* Cambridge: Cambridge University Press.

[4] Chafe, W. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

[5] Deterding, D. 2006. The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association* 36 (2), 87-196.

[6] Eckert, P. 2004. *Vowel Shifts in Northern California and the Detroit Suburbs: Northern California Vowels.* http://web.stanford.edu/~eckert/vowels.html

[7] Fairbanks, G. 1966. *Experimental phonetics: Selected articles.* Urbana: Univ. of Illinois Press.

[8] Gordon, M. 2000. Phonological Correlates of Ethnic Identity: Evidence of Divergence? American Speech 75(2), 115-136.

[9] Hall-Lew, L. 2009. *Ethnicity and Phonetic Variation in a San Francisco Neighborhood.* Ph.D. diss., Stanford University.

[10] Hall-Lew, L. 2011. The Completion of a Sound Change in California English. *Publication of the 17th International Conference of the Phonetic Sciences (ICPhS XVII).* Hong Kong.

[11] Hall-Lew, L. 2013. 'Flip-flop' and mergers-in-progress. *English Language and Linguistics*, 17(2): 359-390.

[12] Hinton, L., Moonwomon, B., Bremner, S., Luthin, H., Van Clay, M., Lerner, J., Corcoran, H. 1987. It's Not Just the Valley Girls: A Study of California English. *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society 13,* 117–28.

[13] Krug, M., Schlüter, J. 2013. *Research Methods in Language Variation and Change.* Cambridge: Cambridge University Press.

[14] Labov, W. 2001. *Principles of Linguistic Change. Volume 2: Social Factors. Language in Society.* Oxford: Blackwell.

[15] Mallinson, C., Childs, B., van Herk, G. 2013. *Data Collection in Sociolinguistics: Methods and Applications.* New York, NY and Abingdon, Oxon: Routledge.

[16] Max Planck Institute for Psycholinguistics, The Language Archives, Nijmegen, The Netherlands. 2011. *ELAN* [Computer program]. Version 4.0.1. URL http://tla.mpi.nl/tools/tla-tools/elan/. Retrieved March 1, 2011.

[17] Moonwomon, B. 1991. *Sound Change in San Francisco English.* PhD dissertation, University of California, Berkeley.

[18] Podesva, R. J. 2007. Phonation Type As A Stylistic Variable: The Use of Falsetto in Constructing A Persona. *Journal of Sociolinguistics* 11.478-504.

[19] Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H. and Yuan, J. 2014. *FAVE 1.1.3.* ZENODO. doi:10.5281/zenodo.9846.

[20] Scarborough, R., Brenier, J., Zhao, J., Hall-Lew, L., and Dmitrieva, O. 2007. An Acoustic Study of Real and Imagined Foreigner-Directed Speech. *Publication of the 16th International Conference of the Phonetic Sciences (ICPhS XVI).* Saarbrücken, Germany.

[21] Sharma, D. 2011. Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, 15(4):464–492.

[22] Troiani, V., Fernández-Seara, M. A., Wang, Z., Detre, J. A., Ash, S. and Grossman, M. 2008. Narrative speech production: an fMRI study using continuous arterial spin labeling. *Neuroimage*, *40*(2): 932–939.

[23] Varon, S. 2007. Robot Dreams. New York: First Second.

[24] Weinberger, S. H. and Kunath, S. A. 2011. The Speech Accent Archive: Towards a typology of English accents. *Language and Computers* 73: 265–281.

---

[i] Author order is alphabetical.