Sociophonetic perspectives on stylistic diversity in speech research

Sociophonetic Perspectives on Stylistic Diversity in Speech Research

[Authors anonymized]

Abstract

Sociolinguistic data collection traditionally includes interviews, reading passages, and word lists (Labov 1972). Researchers have increasingly sought out elicitations tasks that have the benefits of read tasks (studying infrequently occurring variables; controlling for linguistic factors) while also eliciting styles more comparable to interview speech (see, e.g., Drager 2018). Examples include the Map Task (Brown et al. 1984) and the Diapix task (Baker & Hazan 2011). Other researchers have turned instead to elicitations tasks that maximize ecological validity, taking themselves out of the recording context altogether and training participants to collect their own field recordings (see, e.g., Podesva 2007; Sharma 2011). How comparable is the speech elicited from each of these tasks? Building on our previous results (Authors et al. 2015; Authors 2017) we consider three US English speakers' vowel productions from interviews, reading passages, controlled 'Lab Tasks', and self-recordings. Our results suggest fewer differences across tasks than might be expected, suggesting that the interactional context may be more predictive of style shifting than the task, itself (Levon 2013).

1. Introduction

An important aspect of sociophonetic methodology is the context in which data are collected. While this has been foundational to our field since at least Labov (1966), there is today an ever-increasing proliferation of social contexts resulting in speech data being analyzed by (socio)phoneticians, and we are largely lacking a close analysis of the differences that might obtain as a result. The present paper builds on the results from Authors et al. (2015), a vowel analysis of one speaker across multiple sociolinguistic and phonetic speech elicitation tasks, and Authors (2017), a sibilant analysis of an additional three speakers in both interviews and self-recordings (see Podesva 2007; Sharma 2011). Here we present a vowel analysis of those same three speakers across all of these tasks. Given the social and individual differences between these three speakers (despite all being highly educated American women in their mid-20s), the comparisons are strictly intraspeaker, examining the effect of speech elicitation method. One of the primary examinations is on speech elicited using what we call *Lab Tasks*, which here include four non-/minimally scripted laboratory-based speech elicitation tasks: the Map Task (Brown et al. 1984, Anderson et al. 1991), the Diapix task (Baker & Hazan 2011, Tuomainen & Hazan 2018), the narration of a silent film (Chafe 1980), and the narration of a picture book (e.g., Troiani et al. 2008; Author 2018a). In the present paper we ask: when produced in the same sociolinguistic fieldwork context, how does speech obtained from Lab Tasks compare to speech obtained from classic sociolinguistic interview tasks? We compare that contrast to the classic

contrast between interview speech and read speech, and the relatively more innovative contrast between interview speech and self-recorded speech.

The results suggest that, all things being equal, Lab Tasks do elicit speech production patterns that are generally comparable to speech elicited from sociolinguistic interviews. The question remains if the comparability would be maintained if the lab tasks were conducted in a laboratory context. Results also suggest that, in contrast to our previous findings (Authors et al. 2015; Authors 2017), self-recorded speech and interview speech are largely comparable for these three speakers, at the vocalic level, although the self-recorded speech does show a wider range of variation. Furthermore, despite extensive previous research suggesting differences between interview speech and read speech (Labov 1966), we see substantially fewer differences that expected, possibly due in part to the speakers' academic professions. Overall, these results demonstrate the value of a uniquely sociophonetic perspective on data collection, addressing ongoing discussions of speech recording contexts in both variationist sociolinguistics (Mallinson et al. 2013) and phonetics (e.g., Xu 2010; Wagner et al. 2015).

2. Theoretical Framework

The present study is motivated by the desire to better understand how speech elicitation methods not typically seen in sociolinguistics compare to more standard and traditional sociolinguistic fieldwork methods. In our examination we explore the validity of incorporating Lab Tasks into sociolinguistic fieldwork, outside of the highly-controlled speech laboratory setting in which we typically see them implemented. This is an important aspect of the present research and one which we will reiterate throughout this paper: the Lab Tasks, as employed here, should not necessarily be assumed to be comparable to Lab Tasks as implemented in much of the phonetics literature. While the tasks themselves remain the same, the context in which they were recorded is entirely different, with the speakers of the present study being recorded in casual fieldwork-based sociolinguistic interview locations, outwith a speech laboratory setting. This difference is something Wagner et al. (2015) refer to as the "habitat" of the speech act. If we see specific style-shifting patterns within one habitat (e.g., speech from Lab Tasks obtained in a highly-controlled laboratory setting) we must be hesitant to assume that these findings can be shared across other habitats (e.g., speech from Lab Tasks obtained in a sociolinguistic interview setting). That said, it is an empirical question to what extent the task itself, regardless of "habitat," may elicit a distinct speech style. Here we consider the possibility that lab tasks may present new register demands (Silverstein 2003) that must be understood empirically.

Throughout the remainder of this section we provide an overview of the speech elicitation methods implemented throughout the present study. Given the prevalence of the traditional sociolinguistic interview within the field, it seems only logical to begin our discussion there.

The sociolinguistic interview was devised as a way to identify and elicit the speaker's least self-conscious speech style, which was considered the central object of study in early variationist research. This motivation is captured in Labov's Vernacular Principle, which states that: "The style which is most regular in its structure and in its relation to the evolution of language is the vernacular, in which the minimum attention is paid to speech" (Labov 1972: 112). Labov's (1972) model of stylistic variation was based on the principle that a speaker's stylistic repertoire could be placed on a formal-informal continuum, reflecting the amount of attention paid to speech – the more a speaker monitors their production, the more likely they are to adopt prestigious variants. This yields the prediction that the conversational portion of the sociolinguistic interview will elicit the least conservative variants, followed by the reading passage and word list, with minimal pairs likely to result in the most conservative productions. While these predictions were borne out in early work (e.g. Labov 1966), the efficacy of the sociolinguistic interview in capturing stylistic variation has received extensive criticism.

It is now widely recognized that speakers may style shift in response to a number of factors for example, the actual or expected speech of the addressee (Bell 1984) or the topic (Rickford & McNair Knox 1994; Author 2018). More recent conceptualizations of style acknowledge that style-shifting does not occur only in response to external factors, but may reflect speakers' agentive pursuit of social goals -- the desire to project a particular stance or persona; to express alignment or distance with regard to social groups or norms, or to signal a particular relationship between themselves and their interlocutors (e.g. Schilling-Estes 2002).

Despite the extensive theoretical criticism, sociolinguistic interviews remain a staple of variationist research, and with good reason. Although a number of alternative/complementary methods have been adopted (such as self-recordings; see below), the need to collect data efficiently and elicit tokens of the specific variables under study means that the interview and reading tasks are usually the researcher's first choice when designing a sociolinguistic study. Sociolinguists' continuing reliance on the Labovian sociolinguistic interview represents the need to balance practicality with two competing aims: the need to capture a wide range of stylistic variation in the informant's speech, and the need to collect controlled data on the features under study. We propose that sociolinguists may attempt to address this issue by implementing speech elicitation methods typically seen in phonetics and phonology research.

2. 1 The Lab Tasks

The Lab Tasks are speech elicitation methods typically used in phonetics research, which are not yet considered a part of the typical sociolinguistic toolkit. The Lab Tasks used within the present paper are the Map Task (Brown et al. 1984, Anderson et al. 1991), the Diapix Task (Baker & Hazan 2011), a picture book narration (e.g., Troiani et al. 2008; Author 2018a), and a silent film narration (e.g., Chafe 1980).

The Map Task (Brown et al. 1984, Anderson et al. 1991) is historically a diologic task, where both interlocutors are presented with a graphic representation of various locations. These locations and their placement are identical across both interlocutor's 'maps'. One interlocutor has a pre-drawn route around the various locales (the "Instruction Giver"; Anderson et al. 1991: 352) where the other is just given pictures of the locations (the "Instruction Follower") with no path connecting them. The task itself involves the 'Giver' providing instructions to the 'Follower' where the goal is to recreate the original route as shown on the 'Giver's' map. While the labels of 'Giver' and 'Follower' may imply that the 'Giver' leads the conversation, there are no restrictions on how each interlocutor may navigate this task. The Diapix Task (Baker and Hazan 2011) is a 'spot the difference' game where both interlocutors are given a similar, but different picture or scene. The participants only see their version and must discuss the scene to determine where differences lie. In the present analysis, the first author acted as interlocutor to the participants for each of these tasks.[1] Both the Map Task and Diapix are designed to elicit spontaneous speech. Where the Map Task aims to provide the researcher a greater deal of control over the speaking context, the Diapix Task aims to give researchers greater control of the lexical content of the dialogue.

The picture book narration (e.g., Troiani et al. 2008; Author 2018a), and silent film narration (e.g., Chafe 1980) are purely monologic tasks, and relatively self-explanatory in name. For the picture book narration, participants are asked to "read" aloud one chapter from a wordless picture book, *Robot Dreams* (Varon 2007). For the current study, they were not given the chance to look through the book prior to completing the task. Similarly, participants were asked to watch and describe the plot of a short silent film (*The Pear Story*; Chafe 1980) as they were watching it. These tasks provide insights into the construction and organisation of narrative speech (e.g., Bamberg and Marchman 1991), but are further suited quite well for sociolinguistic work as they provide a way for eliciting context-controlled semi-structured non-read speech.

2.2 Self-Recordings

Self-recordings, much like the Lab Tasks, are not as yet typical in sociolinguistic speech elicitation and data collection. Self-recordings, as implemented in contemporary sociolinguistic studies, alter the role of the participant to that of being a temporary fieldworker. In this, the participant is trained in field recording procedures (e.g., microphones, recorders, ethical consent), and is instructed to record themselves as they go about their day-to-day life. In this, self-recorded data may provide insights into individual speaker variation related to the context of the self-recording (e.g., Podesva 2007; 2011a,b), the participants' interlocutors (e.g., Sharma 2011), or any combination of different recording contexts with a range of interlocutors (e.g., Authors 2017). Levon (2013) suggests that the nature of the interactional situation, and not the

[1] Data was also collected not analyzed here (due to too-low token frequencies) from the participants completing both a monologic Map Task (Scarborough, et al., 2007) and a monologic Diapix, without an interlocutor.

fact of the self-recording itself, is what is likely to motivate a stylistic difference relative to an interview, and we suggest (Authors 2017) that this potential for expanding the range of interactions recorded is still precisely what makes self-recordings valuable.

Self-recordings are increasingly being implemented in studies of language variation and several recent studies have found a significant difference between interview speech and self-recorded speech with respect to a wide range of linguistic variables (Tseng 2014; Van Hofwegen 2016; Authors 2017), although not always (Saisuwan 2016). Furthermore, although the elicitation of a wider stylistic range provides better empirical evidence of a speaker's possible range of variation, this evidence comes at the expense of the researcher understanding the full social context in which the recording was made, and therefore all possible causal factors. Analyses such as ours are in service of understanding if self-recorded data is sufficiently insightful to justify the costs, which can be substantial for both the researcher and the participant-turned-fieldworker.

2.3 The Current Study

The present paper builds upon the findings presented in Authors et al. (2015) and Authors (2017). Authors et al. (2015) presents an analysis of five vowel changes associated with the California Vowel Shift (Eckert 2004) produced by a single female speaker from San Francisco, California, examining vowel productions across sociolinguistic interviews, reading passages, Lab Tasks, and self-recordings. Similar to the present paper, all stylistic differences are viewed in relation to interview speech productions. For the most part, Lab Task speech was shown to be largely similar to interview speech, though there is a great deal of variability within and between the individual Lab Tasks. What stood out was a reliable difference between self-recordings and interview speech wherein self-recorded data was shown to elicit more advanced vowel productions in terms of the California Vowel Shift. The (2015) results also suggested surprisingly little difference between read-speech and interview speech across the examined vowels.

Expanding on these results, Authors (2017) explore differences in /s/ production between read speech, self-recorded speech, and interview speech for four speakers: the same speaker from Authors et al. (2015) and three other speakers, who are the same three speakers of the present study: 'Kat', 'Piper', and 'Virginia' (Table 1). Authors (2017) show a consistent and reliable effect of /s/ being produced with a higher center of gravity (CoG) when occurring in read speech compared to interview speech. This effect is rather unsurprising given previous literature on the relationship between 'clear-speech' and 'conversational speech' in sibilant variation (e.g., Maniwa et al. 2009; Tucker et al. 2016). Stylistic differences can also be seen between self-recorded speech and interview speech for all speakers' /s/ productions. This effect is often equal to, or even greater than, the results seen between read-speech and interview speech, but is variable in the directionality of difference based on speaker and self-recorded context. In other

words, there is no consistent pattern which will predict if a speaker will produce /s/ with a higher or lower CoG in the self-recordings compared to sociolinguistic interviews, but that differences between self-recordings and interviews are mediated by the recording context of the self-recorded session itself (i.e., one-on-one conversation, group hang out, Skype conversation, etc.) and are highly subject to variability by speaker. Even though no consistent patterns are seen for the /s/ data, the self-recordings provided the opportunity to capture a wider range of variation than would have been seen via sociolinguistic interviews alone.

3. Methods

Data for the present study comes from speech from three highly educated, cis-gendered females of varying ethnicities, born in the United States between 1985 and 1990 (Table 1). Each speaker took part in a sociolinguistic interview which included multiple reading passages and Lab Tasks. All speakers then completed several self-recordings in the weeks following the original recording sessions (see Authors 2017).

All recordings were made in summer 2016, in Edinburgh, Scotland, where the researchers and participants were all working and studying at the time. Each recording took place in the participants' homes, in a single recording session. All interviews were conducted by the first author, who, like the speakers, is an American female academic living in Edinburgh.[2] Recording sessions typically lasted between 3-4 hours. While these are relatively long recording sessions, all speakers are friends with one of the two authors, and recording sessions were more casual than a typical interview context, with frequent breaks and chit-chat between tasks, and with frequent checks to avoid participant fatigue.

	Born	Hometown	Heritage	US Gen	Work
Kat	1986	SF Bay Area, CA	Chinese	2	Assistant Professor
Piper	1988	Louisville, KY	Greek	3	PhD Student
Virginia	1990	San Antonio, TX	Mexican	3	PhD Student

Table 1: Speakers and Social Characteristics. 'SF' = San Francisco; US Gen = Immigration Generation to the US.

As discussed in the previous section, the Lab Tasks utilized for the present study are: Map Task (Brown et al. 1984, Anderson et al. 1991), the Diapix Task (Baker & Hazan 2011, Tuomainen & Hazan 2018), a picture book narration (e.g., Troiani et al. 2008; Author 2018a), and a silent film narration (e.g., Chafe 1980). We also analyze the following reading passages: *The Rainbow*

[2] Two of Virginia's self-recordings were made with the second author present, an American male academic living in Edinburgh.

Passage (Fairbanks 1966), *The Boy Who Cried Wolf* and *North Wind and the Sun* (Deterding 2006), *Please Call Stella* (Weinberger and Kunath 2011), and three from Gordon (2000): *Basketball*, *Dolls*, and *Victoria's Secret*. In the mixed effects models presented below, the specific Lab Task or specific Reading Passage is treated as a random effect; the tokens are too few to examine the effect of specific Task/Passage as a main effect.

All audio was transcribed in ELAN (ELAN 2017; Wittenburg et al. 2006) and underwent phone forced-alignment using FAVE (Rosenfelder et al. 2015). Segments of speech containing external background noise or overlapping speech were excluded from analysis. Single-point first and second formant values from ten vowels (BAT, BAIT, BET, BEET, BIT, BOUGHT, BUT, BOAT, BOOT, & BOT) were obtained via FAVE-extract (Rosenfelder et al. 2015), using the default measurement points of that script. Formant values were normalized (Lobanov 1971), although each speaker's productions are analyzed separately. Vowel tokens preceded or followed by a non-phonemic noise (e.g., cough, laugh) or another vowel were excluded from the analysis.

We ran 180 model comparisons (10 vowels, 2 vowel formants, 3 speakers, and 3 task comparisons). Statistical significance was determined by fitting separate standard mixed effects models (*lme4*, version 1.1-15) for each speaker and each normalized formant (F1, F2) of each vowel. Best-fit models were determined by step-wise, drop-one model comparisons. Fixed effects included vowel duration, preceding and following phonological environment (with separate effects for manner, place, and voicing), and elicitation task. Word was included as a random intercept in all models and an additional random intercept was included depending on the task comparison of the model.[3] For example, initial models considering self-recordings included a random intercept for the specific self-recording, which was only removed from the best-fit model if indicated by model comparison. Elicitation task was always a binary factor, with separate comparisons between interviews and reading passages, interviews and lab tasks, and interviews and self-recordings. Because the nature of the study is exploratory and the null hypothesis of interest, we adjusted the alpha value for significance to 0.0008, using a basic Bonferroni correction.

4. Results

Out of all 180 models, only 23 (13%) models showed a significant effect of elicitation task at the adjusted *p*-value.[4] For all speakers, the difference between interviews and lab tasks was much less than either the difference between interviews and read speech or between interviews and self-recordings. For two speakers, the comparisons with read speech resulted in slightly more differences than the comparisons with self-recordings; for the third speaker, the number of

[3] Example model: f1_Kat_IvwRdg_BEET <- lmer(f1 ~ duration + folmanner + folplace + folvoice + precmanner + precplace + precvoice + taskcontrast + (1|word) + (1|reading), data=KatIvwRdgBEET)

[4] Only an additional 16 models showed a significant effect of elicitation task at $p < 0.01$ (for a total of 16+23 = 39 models or 22% of all model comparisons).

differences was the same. Only one speaker showed any significant difference between interviews and lab tasks: only one formant value, for one vowel. Overall, the results for these three speakers suggest that lab tasks generally elicit the same speech style as sociolinguistic interviews, all else being equal, whereas the difference between self-recorded speech and interview speech is quantitatively comparable to the well-known differences between read speech and interview speech.

4.1 Interview speech vs. Read speech

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	F1	F2	F2	F2	--	--	--	--	--
Piper	--	--	F1	--	--	--	--	F1	F1	--
Virginia	F1	--	F1	--	--	--	--	--	--	F1, F2

Table 2: All models indicating a significant difference between interview speech and reading passage speech at $p < 0.0008$, with a random effect of specific reading passage. All values show a positive effect (larger formant value) for read speech.

Given the well-known phonetic differences between speech elicited in interviews and speech elicited from reading passages, the results show fewer vowel contrasts than one might expect. Kat and Virginia both show only four of 20 possible effects of task, and Piper only three.[5] This low number may be due to all participants being professional academics, a profession that may favor a more 'read' style of spontaneous speech in interview contexts (e.g., because academics have to verbally perform expertise on a regular basis), and/or a more 'casual' style in read speech contexts (e.g., because academics often read aloud for teaching or presentation purposes).[6] The low number of contrasts may additionally arise from the particularly casual and friendly interactional style of the recording context, which would promote a reduction in style shifting.

Across the differences that do obtain there is a fairly even spread between F1 and F2 effects, although with some speaker-specific differences; Piper and Virginia favor F1 contrasts while Kat favors F2 contrasts. In all cases an F1 effect indicates a lower vowel in read speech than interview speech, and an F2 effect indicates a fronter vowel in read speech than interview speech. All three speakers appear to be using a more enunciated or 'clear speech' style in the read tasks (see Smiljanić & Bradlow 2009) which is obtained by the use of a more expanded vowel space. Kat's BAT F2 effect does not indicate a more expanded vowel space, but may be the results of a style shift away from the backed variant, which associated with a certain stigmatized Californian persona. We will return to this idea further on.

[5] Seven additional models showed an effect of read versus interview speech at alpha level 0.01: Kat F1 BAT & F2 BEET, BET, & BOT; Piper F1 BET & BOUGHT; Virginia F1 BIT.
[6] Note that the speaker in Authors et al. (2015) was not an academic but was in law school at time of recording.

4.2 Interview speech vs. Lab Task speech

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	--	--	--	F2	--	--	--	--	--
Piper	--	--	--	--	--	--	--	--	--	--
Virginia	--	--	--	--	--	--	--	--	--	--

Table 3: All models indicating a significant difference between interview speech and lab task speech at $p < 0.0008$, with a random effect of specific lab task. All values show a positive effect (larger formant value) for read speech.

The most striking result is that 59 of 60 models showed no significant effect of the task difference between interview and lab tasks.[7] The one exception was Kat's production of the BAT vowel in F2 space.[8] We suggest that the lack of differences between speech elicited from an interview and speech elicited from tasks typically associated with the laboratory may be precisely the fact that the elicitation took place in a relaxed interview context and not in a laboratory context. Lab tasks were treated by both interviewer and interviewee as either fun games (Map Task, Diapix) or an extension of the interview: describing a silent film and chuckling about its content is not too dissimilar from describing a past memory and chuckling about *its* content. We set up this contrast in the most careful and conservative way possible, to keep all things equal and only vary the use of lab tasks versus the use of interview questions. Given this context, it appears that there is little about the tasks *themselves* that elicits a change in speech style. It remains to be seen what the effect is between speech obtained in a fieldwork context and speech obtained in a laboratory context, all else being equal.

Although only one difference obtains between interview speech and lab task speech, it is a highly significant difference, and one that may have a stylistic explanation. Kat produces a fronter production of the BAT vowel in the lab tasks than in the interview. It is sociolinguistically relevant that it is BAT F2, specifically, which is varying. Kat is from the San Francisco Bay Area, where BAT backing has been documented a change in progress (Cardoso et al. 2016; Hall-Lew et al. 2015). The backed production has been seen to index a range of social meanings (D'Onofrio 2015, 2018), but particularly meanings associated with the stigmatized Californian persona, 'Valley Girl' (see D'Onofrio 2018). Kat, as a young, female Californian, may vary her production of BAT depending on her motivation to distance herself from these meanings. The previous literature suggests that fronting BAT may be one of the more effective ways of doing this. This would explain why Kat produces backed BAT in her interview task, as expected, but

[7] This result contrasts with our preliminary results as presented at NWAV47 (Authors 2018b). Those results defined the phonological context structure differently and did not model a random effect of specific lab task.

[8] Even at $p < 0.01$, only three additional variables show a significant difference between interview speech and lab task speech: Kat BOOT F1, Piper BOAT F2, and Virginia BOT F1.

fronts the vowel in both read speech and the lab tasks. Although everything else remains the same in the recording session, Kat explicitly orients differently to the lab tasks, saying: "I know it's not the point of the task, but I wanna win!" For Kat, this competitive drive for excellence may characterize a shift in register demand (Silverstein 2003), invoking a style shift that results in the reduction of at least one feature that indexes a 'Valley Girl' persona or its related stigmatized meanings.[9] Although there is a small amount of read speech in the lab tasks (the labels on the Map Task maps and occasional labels on the DiaPix pictures), Kat's BAT F2 shift is not straightforwardly explained by this. Her picture book narration and Map Tasks show her frontest articulations, and her silent movie narration and DiaPix speech show her backest.

In summary, while the results overall indicate that the elicitation tasks typically used in laboratory contexts do not, in an interview context, elicit speech that is significantly different from interview speech, there may be exceptions. Furthermore, those exceptions may be sociolinguistically relevant. With respect to accurate phonetic description, an analysis of Kat's vowel production that had been based only on the speech from lab tasks would have represented her as a less advanced speaker of this particular regional sound change than would an analysis based on speech obtained from an interview, even when the lab tasks were performed by familiar interviewer and in her own home.

4.3 Interview speech vs. Self Recorded speech

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	--	--	*-F1*	--	--	--	--	--	*-F1*
Piper	--	--	F1	F1	--	F1	--	F1	--	F1
Virginia	--	--	--	--	--	--	--	--	--	F1

Table 4: All models indicating a significant difference between interview speech and self-recorded speech at $p < 0.0008$, with a random effect of specific self-recording. Negative values and italics indicate a smaller formant value for self-recorded speech.

In methodological contrast to lab task elicitation is elicitation based on self-recordings. Here we again see relatively few differences overall: eight out of 60 model comparisons showed a significant difference between interview speech and self-recorded speech.[10] We also see a wide difference between speakers, with only one effect for Virginia, two for Kat, and five for Piper. The most striking result here is that all of the tasks effects pertain to variation in F1, not F2, although one speaker (Kat) shows higher vowels in self-recorded speech, while the other two

[9] This could be investigated in future research by examining other features that index 'Valley Girl' or other equally stigmatized but more contemporary Californian personae.

[10] Six additional models were significant at $p < 0.01$: Kat F1 BOAT & F2 BEET; Piper F1 BOOT & BET; Virginia F1 BAIT & BEET.

show lower vowels in self-recorded speech. This is somewhat reminiscent of the results from Authors (2017), wherein Kat showed a higher (fronter) /s/ center of gravity (CoG) in self-recordings, and Piper and Virginia showed lower (backer) CoG values in self-recordings, relative to interview speech. This suggests that Kat may be treating self-recordings stylistically differently than the other two speakers, perhaps in part because of the presence of the first author. Unlike the interview, reading passage, and lab tasks, for all three speakers, most of the self-recordings were made without the first author present. The exception was one of Kat's three self-recordings, where the first author was present for a casual lunch with a third mutual friend. It might also be the case that all interactions with the first author are more similar to a self-recording than for the other two speakers (who met the first author in the research context, via an introduction from the second author). However, qualitative comparisons show that the difference between Kat's interview speech and self-recorded speech is greater than the differences between Kat's speech towards different interlocutors. Similarly, two of Virginia's three self-recordings were made with the second author present, but qualitative analysis shows no differences between her three self-recordings.

5. Conclusion

This paper has considered the role of speech elicitation method with respect to phonetic and sociolinguistic methodology. In comparing vowel variation for three speakers across four different elicitation techniques, we find what are perhaps surprisingly few differences.

Our results do show several differences between read-speech and interview speech, which itself is not surprising. However, there are fewer differences than one might expect given the well-known differences between this "clear-speech" style and that of the more "conversational" style associated with casual sociolinguistic interviews. This may be due in part to the fact that all three speakers are professional academics whose read and conversational speaking styles are perhaps more similar to one another than for a typical individual.

Results further suggest that Lab Tasks are a valuable resource for implementation into sociolinguistic fieldwork methods, showing surprisingly few differences to interview speech. As such, Lab Tasks offer sociolinguists a wider range of elicitation methods to draw from and allows for a greater degree of control for any speech variable under consideration. In this the Lab Tasks may highlight specific sociolinguistically relevant variables employed by any given speaker. Indeed, our own results suggest that where difference does arise (i.e., Kat with a more fronted BAT vowel in Lab Task speech), the local indexical meaning of the variant may be a driving factor. In this regard, reliance only on Lab Tasks as representative of speech production patterns risks potentially misrepresenting a speaker's typical production of those phonetic variables which are most indexically rich. Further, despite seeing very few differences overall between Lab Task speech and interview speech, we want to caution against comparing these

results to those seen in the wider phonetics literature where these tasks are often utilized, because of the interactional differences between our study and studies that obtained their data in a more typical laboratory context.

Results from the self-recorded speech dovetail with our previous work (Authors et al 2015; Authors 2017) revealing a greater degree of variability than can be captured through traditional sociolinguistic interviews alone. Though self-recorded data inherently entails multiple ethical and logistical challenges, these results, in combination with those of other studies examining self-recorded data such as Sharma (2011) and Podesva (2007), highlight the value of incorporating self-recorded data into sociolinguistic research despite the challenges that they bring.

Wagner et al. suggest that "if our methods limit the way data is collected (i.e., the way speech is produced) the explanatory power of the results is also limited" (2015: 4). Incorporation of self-recordings and Lab Tasks into the sociolinguists toolkit provides us with two separate but equally valuable implications for sociolinguistic methodological practice: through self-recordings there exists the potential for a wider range of phonetic variability on the one hand, and on the other, Lab Tasks allow for more control over the speech elicited from participants. As such, we suggest that both have a place within the wider sociolinguistic methodological practice.

References

Author (2018a)
Authors (2018b)
Authors (2017)
Authors et al. (2015)
Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson and Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech*. 34(4): 351—366.
Baker, Rachel, and Valerie Hazan. 2011. DiapixUK: A task for the elicitation of spontaneous speech dialogs. *Behavior Research Methods*, 43(4), 761-770.
Bamberg, Michael, and Virginia Marchman. 1991. Binding and unfolding: Towards the linguistic construction of narrative discourse. *Discourse Processes*, *14*(3), 277-305.
Bell, Allan. 1984. Language style as audience design. *Language in Society*, 13, 145-204.
Brown, G., Anderson, A. H., Shillcock, R. and Yule, G. 1984. *Teaching Talk: Strategies for production and assessment.* Cambridge: Cambridge University Press.
Cardoso, Amanda, Lauren Hall-Lew, Yova Kemenchedjieva, and Ruaridh Purse. 2016. Between California and the Pacific Northwest: The Front Lax Vowels in San Francisco English. In Valerie Fridland, Betsy Evans, Tyler Kendall, and Alicia Wassink, eds. *Speech in the*

Western States, Volume 1: The Coastal States, pp. 33-54. Publication of the American Dialect Society. Durham, NC: Duke University Press.

Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

D'Onofrio, Annette. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, *19*(2), 241–256.

D'Onofrio, Annette. 2018. Controlled and automatic perceptions of a sociolinguistic marker. *Language Variation and Change*, 30(2), 261-285.

Deterding, David. 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), 187–196.

Drager, Katie. 2018. *Experimental Research Methods in Sociolinguistics*. Bloomsbury Publishing.

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

ELAN (Version 5.0.0-beta). (2017). Nijmegen: Max Planck Institute for Psycholinguistics.

Fairbanks, Grant. 1966. *Experimental phonetics: selected articles*. University of Illinois Press.

Gordon, M., Barthmaier, P., & Sands, K. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174.

Hall-Lew, Lauren and Amanda Cardoso, Yova Kemenchedjieva, Kieran Wilson, Ruaridh Purse, and Julie Saigusa. 2015. San Francisco English and the California Vowel Shift. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 591.

Labov, William. 1966. *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington DC.

Labov, William. 1972. Some Principles of Linguistic Methodology. *Language in Society*, *1*(1), 97–120.

Levon, Erez. 2013. Ethnography and recording interaction. In *Research Methods in Linguistics*, R.J. Podesva and D. Sharma, eds. 195–215. Cambridge: Cambridge University Press.

Lobanov, Boris M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America.* 49(2B):606–608.

Mallinson, Christine, Becky Childs, and Gerard Van Herk. 2013. *Data collection in sociolinguistics: Methods and applications*. Routledge.

Maniwa, Kazumi, Allard Jongman, & Travis Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America* 125:3962–3973.

Podesva, Robert J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11:478–504.

Podesva, Robert J. 2011a. Salience and the Social Meaning of Declarative Contours: Three Case Studies of Gay Professionals. J*ournal of English Linguistics* 39:233–264.

Podesva, Robert J. 2011b. The California Vowel Shift and Gay Identity. *American Speech* 86:32–51.

Rickford, John R. and Faye McNair-Knox. 1994. Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. *Sociolinguistic perspectives on register*, pp.235-276.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2015. FAVE (Forced Alignment and Vowel Extraction) 1.1.3. ZENODO. doi:10.5281/zenodo.9846.

Saisuwan, Pavadee. 2016. Male Femininity in Thai Among Men who Identify with Non-Normative Male Roles. Doctoral dissertation, Queen Mary University of London.

Scarborough, Rebecca, Jason Brenier, Yuan Zhao, Lauren Hall-Lew, and Olga Dmitrieva. 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.

Schilling-Estes, Natalie. 2002. Investigating stylistic variation. In J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 375-401). Malden: Blackwell.

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

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3-4), 193-229.

Smiljanić, Rajka, and Ann R. Bradlow. 2009. Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass* 3(1): 236–264.

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.

Tseng, Amelia. 2015. Vowel Variation, Style, and Identity Construction in the English of Latinos in Washington, DC. Doctoral dissertation, Georgetown University.

Tucker, Benjamin, Viktor Kharlamov, and Daniel Brenner. 2016. What's the Zed? The Acoustics of Conversational Fricatives in Mid-Western American English. Paper presented at NorthWest Phonetics & Phonology Conference 2016. University of Oregon, Eugene, Oregon.

Tuomainen, Outi, and Valerie Hazan. 2018. Investigating Clear Speech Adaptations in Spontaneous Speech Produced in Communicative Settings. In: *Challenges in Analysis and Processing of Spontaneous Speech,* pp. 9-25. Research Institute for Linguistics of the Hungarian Academy of Sciences. Nyelvtudományi Intézet, Budapest.

Van Hofwegen, Janneke. 2016. A day in the life: What self-recordings reveal about "everyday" language. Paper presented at NWAV 45, Vancouver, BC.

Varon, Sara. 2007. *Robot Dreams*. New York: First Second.

Wagner, Petra, Jürgen Trouvain, and Frank Zimmerer. 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics* 48: 1-12.

Weinberger, S. H., & Kunath, S. A. 2011. The Speech Accent Archive: Towards a typology of accents. *Language and Computers-Studies in Practical Linguistics*, *73*(1), 265-281.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, 1556–1559.

Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336.

Sociophonetic Perspectives on Stylistic Diversity in Speech Research
[Authors anonymized]

Abstract

Sociolinguistic data collection traditionally includes interviews, reading passages, and word lists (Labov 1972). Researchers have increasingly sought out elicitation tasks that have the benefits of read tasks (studying infrequently occurring variables; controlling for linguistic factors; eliciting the same lexical items across participants) while also eliciting styles more comparable to interview speech (see Drager 2018). Examples include the Map Task (Brown et al. 1984) and the Diapix task (Baker and Hazan 2011). Other researchers have turned instead to elicitation tasks that maximize ecological validity, taking themselves out of the recording context and training participants to collect field recordings (e.g., Podesva 2007; Sharma 2011). How comparable is the speech elicited from each of these tasks? Building on previous results (Authors et al. 2015; Authors 2017) we consider three US English speakers' vowel productions from interviews, reading passages, 'Lab Tasks', and self-recordings. We suggest that both Lab Tasks and self-recordings have the potential to increase descriptive accuracy and indexical analysis in sociophonetic research.

1 Introduction

A key aspect of sociophonetic methodology is the task used to elicit speech. While some task-based variation is foundational to the field (Labov 1966), we have less knowledge about the expected effects of other tasks. This paper builds on Authors et al. (2015), a vowel analysis of one speaker across multiple speech elicitation tasks, and Authors (2017), a sibilant analysis of an additional three speakers in interviews and self-recordings (see Podesva 2007; Sharma 2011). Here we present a vowel analysis of those same three speakers across all tasks, considering an expanded range of vowels, to see if the results for one speaker (Authors et al. 2015) are comparable for other speakers, and if the patterns for sibilants (Authors 2017) are comparable for vowels. We also focus more in this paper on the speech elicited using *Lab Tasks*. These include three non-/minimally scripted tasks from previous laboratory-based research: Map Task (Brown et al. 1984, Anderson et al. 1991), Diapix (Baker and Hazan 2011, Tuomainen and Hazan 2018), and silent film narration (Chafe 1980), and one from previous field research: picture book narration (e.g., Ravindranath 2008; Stanford 2010; Author 2018a). We ask: when produced in the same fieldwork context, how does Lab Task speech compare to speech from sociolinguistic interview tasks? How do both compare to speech from self-recordings?

Results suggest that Lab Tasks elicit speech production patterns that are roughly comparable to speech elicited from interviews, making them an appealing option for controlled speech elicitation in field contexts, though with one cautionary exception. Results also show that self-

recorded and interview speech are largely comparable with respect to F2, but that F1 varies more, and in different directions between speakers. These results demonstrate the value of a uniquely sociophonetic perspective on data collection, directly testing for potential style-shifting between multiple cases of spontaneous speech. In this, we address ongoing discussions of speech recording contexts in both sociolinguistics (Mallinson et al. 2013; Drager 2018) and phonetics (e.g., Xu 2010; Wagner et al. 2015).

2 Theoretical Framework

This study is motivated by the desire to better understand how speech elicitation methods not typically seen in field-based sociolinguistics compare to more typical fieldwork methods. One appealing option is a 'Lab Task': a method that, e.g., gives the researcher the ability to elicit infrequently occurring variables, to control for linguistic variables, and to elicit the same lexical items across participants. The question is if these tasks are stylistically comparable to interviews, or if they may in fact introduce yet another 'stylistic context' on, e.g., Labov's (1972) attention-based continuum. To test this, we situate Lab Tasks in a sociolinguistic fieldwork setting rather than the highly-controlled laboratory setting where we typically see them implemented. This is an important aspect of the present research: the Lab Task data presented here likely differs from Lab Task data in much of the phonetics literature, because while the tasks remain the same, the recording context is entirely different. Wagner et al. (2015) call this difference the "habitat" of the speech act. If we see style-shifting patterns within one habitat (e.g., Lab Task speech obtained in a lab) we must be cautious in generalizing these findings to other habitats (e.g., Lab Task speech obtained in an interview setting). By controlling for habitat, we consider the possibility that Lab Tasks themselves may present new register demands (Silverstein 2003) that have empirical consequences.

The classic sociolinguistic interview was devised to identify and elicit the speaker's most and least self-conscious speech styles. This motivation is captured in Labov's Vernacular Principle: "The style which is most regular in its structure and in its relation to the evolution of language is the vernacular, in which the minimum attention is paid to speech" (Labov 1972: 112). Labov's model of stylistic variation was based on the principle that a speaker's stylistic repertoire could be placed on a formal-informal continuum, reflecting the amount of attention paid to speech; the more a speaker monitors their production, the more likely they are to adopt ('overtly') prestigious variants. This yields the prediction that the conversational portion of the interview will elicit the least conservative variants, followed by the reading passage, then the word list, and lastly minimal pairs.

Speakers may style shift in response to a number of factors, for example, the actual or expected speech of the addressee (e.g., Bell 1984, Sharma 2011) or the topic (Rickford and McNair Knox 1994; Becker 2014; Author 2018; Grieser 2019). Style-shifting does not occur only in response to external factors, but may reflect speakers' agentive pursuit of social goals: the desire to project

a particular stance or persona, to express alignment or distance with regard to social groups or norms, or to signal a relationship between themselves and their interlocutors (e.g. Schilling-Estes 2002). Because components of the sociolinguistic interview present different register demands (Silverstein 2003), some of these motivations can be evidenced using traditional elicitation methods. For this reason, along with the need to collect data efficiently and elicit tokens of specific variables, interviews remain a staple of variationist research. This paper asks if sociophonetic goals may be even better met by integrating other sociolinguistic methods (like self-recordings), and methods from phonetics and laboratory phonology (like Lab Tasks).

2.1 The Lab Tasks

What we call Lab Tasks are speech elicitation methods typically used in phonetics research, which are not yet considered typical to sociolinguistics: Map Task (Brown et al. 1984, Anderson et al. 1991), Diapix (Baker and Hazan 2011), and silent film narration (e.g., Chafe 1980). We also include a picture book narration (e.g., Ravindranath 2008; Stanford 2010; Author 2018a), because even though it has been implemented in sociolinguistic fieldwork, it is similarly designed to elicit lexically controlled but spontaneous speech.

The Map Task (Brown et al. 1984, Anderson et al. 1991) was designed to elicit collaborative spontaneous speech in an experimental context. Two participants are presented with pictures of various fictional locations on a map. One participant has a pre-drawn route (the "Instruction Giver"; Anderson et al. 1991: 352) and the other (the "Instruction Follower") does not. The Giver provides instructions to the Follower, and the goal is to recreate the route shown on the Giver's map. The Diapix (Baker and Hazan 2011) was similarly devised to give the researcher the "increased ability to generalise results…to naturally occurring speech" (Baker and Hazan 2011: 761), in order to improve upon phonetic analyses that were based only on read-speech (obtained in the lab). Diapix is a 'spot the difference' game where two participants are given a picture and the two pictures differ in a few ways. The participants only see their version and must discuss the scene to determine where differences lie. In the present analysis, the interviewer (the first author) acted as the co-participant for each of these tasks.[1]

The picture book narration (e.g., Ravindranath 2008; Stanford 2010; Author 2018a), and silent film narration (e.g., Chafe 1980) are monologic tasks. For the former, participants narrate from a wordless picture book (*Robot Dreams*; Varon 2007). For the latter, participants describe the plot of a short silent film (*The Pear Story*; Chafe 1980). These tasks provide insights into the construction and organisation of narrative speech (e.g., Bamberg and Marchman 1991), but are

[1] Data was also collected but not analyzed here (due to too-low token frequencies) from the participants completing both a monologic Map Task (Scarborough, et al., 2007) and a monologic Diapix, without an interlocutor. Note that, although there is a very small amount of partially read-speech in the Lab Tasks (the Map location labels or an occasional label in the DiaPix), removing these tokens does not alter the results.

further suited quite well for sociolinguistic work as they provide a way for eliciting context-controlled semi-structured non-read-speech.

2.2 Self-Recordings

Self-recordings, like Lab Tasks, are not yet typical in sociolinguistic speech elicitation. Self-recordings train the participant to also be a fieldworker. The participant learns field recording procedures and records themselves in their day-to-day life. Self-recorded data may provide insights into individual speaker variation related to any number of contextual differences between the recordings (e.g., Podesva 2007; 2011a,b; Sharma 2011; Authors 2017). Levon (2013) suggests that the nature of the interactional situation, and not the fact of the recording being self-recorded, is what is likely to motivate a stylistic difference relative to an interview. Indeed, we argue (Authors 2017) that this potential for expanding the range of interactional situations is precisely what makes self-recordings appealing.

Recent studies have found significant differences between interview speech and self-recorded speech for a range of linguistic variables (Tseng 2014; Van Hofwegen 2016; Authors 2017; but see Saisuwan 2016). Though eliciting a wider stylistic range provides better empirical evidence of a speaker's range of variation, this comes at the expense of the researcher's understanding of the recording context. Our analysis is in service of understanding if self-recorded data is sufficiently insightful to justify the costs, which can be substantial for both the researcher and the participant-turned-fieldworker.

2.3 The Current Study

Previously, Authors et al. (2015) analyzed five vowel changes associated with the California Vowel Shift (henceforth CVS; Eckert 2004) produced by one Chinese American, heterosexual, cis-gendered woman from San Francisco, California, comparing an interview, reading passages, Lab Tasks, and self-recordings. Lab Task speech was largely shown to be similar to interview speech. What stood out was a reliable difference between self-recordings and interview speech: self-recorded data mostly evidenced more advanced productions of the CVS. In the present paper we examine another three speakers, but we expand our vowel analysis to include ten English vowels produced by speakers from disparate geographic regions in the United States.

Authors (2017) explore differences in /s/ production between read-, self-recorded, and interview speech for four speakers: the same speaker from Authors et al. (2015) and three other speakers (the same three speakers of the present study; Table 1). Authors (2017) found a reliable effect of /s/ being produced with a higher center of gravity (CoG) in read-speech compared to interview speech, perhaps similar to the previously documented 'clear-speech' and 'conversational-speech' variation in sibilants (Maniwa et al. 2009; Tucker et al. 2016). Differences found between self-

recorded speech and interview speech were often equal to or greater than the differences between read-speech and interview speech. Furthermore, we found no consistent preference for /s/ CoG productions being higher or lower in the self-recordings; rather, differences appear to be mediated by individual speaker differences and the context of the self-recorded session (i.e., one-on-one conversation, group hangout, Skype conversation). Overall, the self-recordings allowed us to capture a wider range of variation than would have otherwise been observed.

3 Methods

Data come from speech from three highly-educated, heterosexual, cis-gendered women of varying ethnicities, born in the United States between 1985 and 1990 (Table 1). Each speaker took part in a sociolinguistic interview which included multiple reading passages and Lab Tasks, and then completed several self-recordings in the weeks following (see Authors 2017). Interviews took place in 2016 in participants' homes, in a single recording session conducted by the first author, who is older but otherwise demographically similar to them. All speakers are friends with one of the two authors, though only Kat personally knew the interviewer prior to the study. No participant is a sociolinguist or phonetician, and none knew the purpose of the study.

Table 1: Speakers and Social Characteristics. 'SF' = San Francisco; US Gen = Immigration Generation to the US.

Pseudonym	**Born**	**Hometown**	**Heritage**	**US Gen**	**Work**
Kat	1986	SF Bay Area, CA	Chinese	2	Assistant Professor
Piper	1988	Louisville, KY	Greek	3	PhD Student
Virginia	1990	San Antonio, TX	Mexican	3	PhD Student

We obtained readings of seven reading passages: *The Rainbow Passage* (Fairbanks 1966), *The Boy Who Cried Wolf* and *North Wind and the Sun* (Deterding 2006), *Please Call Stella* (Weinberger and Kunath 2011), and three from Gordon (2000): *Basketball*, *Dolls*, and *Victoria's Secret*. The mixed effects models presented below treat task type as a main effect, and for the subset of models with task type as a binary factor, specific Lab Task or specific Reading Passage is further treated as a random effect (tokens are too few to examine them as main effects).

All audio was transcribed in ELAN (ELAN 2017; Wittenburg et al. 2006) and underwent phone alignment using FAVE (Rosenfelder et al. 2015). Single-point first and second formant values from ten vowels (BAT, BAIT, BET, BEET, BIT, BOUGHT, BUT, BOAT, BOOT, and BOT) were obtained via FAVE-extract (Rosenfelder et al. 2015), using FAVE's default measurement points (Labov, et al., 2013). Formant values were z-score normalized (Lobanov 1971). Overlapping speech,

speech containing external background noise, and vowel tokens preceded or followed by a non-phonemic noise or another vowel were excluded from the analysis.

Statistical significance was determined by fitting separate standard mixed effects models (*lme4*, version 1.1-15), with best-fit models determined by hand using step-wise, drop-one, AIC model comparisons. Models for F1 are always separate from models for F2. All model fitting begins with seven fixed effects: vowel duration, and preceding and following phonological environment, with separate effects for manner, place, and voicing. Word is always a random intercept. We explore the effect of task type (interview, Lab Task, reading, self-recording) at three levels of granularity, always with 'interview' set to the reference level. First, we model all the data combined in two models: one for F1, one for F2. In these models, vowel is a fixed effect.[2] Second, we consider the data in 20 separate models, one for each vowel (and formant). In both approaches, speaker is a random intercept.[3] Third, we consider the data in 180 models, one for each vowel, formant, speaker, and binary task comparison (each task type versus interview); for the relevant subsets of these models we included the aforementioned additional random intercept for specific task.[4] Given the multiple comparisons, data are presented at various adjusted alpha levels, as indicated below.[5]

4 Results

4.1 Models From Combined Data

The first results find all task type comparisons to be significantly[6] different from the interview context: Lab Tasks favor a smaller F1 and F2, overall, while read-speech and self-recordings both favor a larger F1 and F2, overall. The variability of both formants is most explained by vowel identity, then by vowel duration, then by task type, followed by all phonological environment factors. The specific results are not presented here, because even when taking vowel identity into account, we are cautious to assign much meaning to them, particularly given the results from the second and third stages of data modelling.

Table 2 shows the results of the second modeling stage.[7] Comparing across all three speakers, task type is a significant model predictor for all 20 models at $p < 0.05$, and for 17 models at $p <$

[2] E.g., f1 ~ vowel + duration + folmanner + folplace + folvoice + precmanner + precplace + precvoice + tasktype + (1|word) + (1|speaker), data=data

[3] E.g., f1 ~ duration + folmanner + folplace + folvoice + precmanner + precplace + precvoice + tasktype + (1|word) + (1|speaker), data=BEET

[4] E.g., f1 ~ duration + folmanner + folplace + folvoice + precmanner + precplace + precvoice + tasktype + (1|word) + (1|reading), data=KatIvwRdgBEET

[5] All 202 model outputs are available by contacting the first author.

[6] All effects show $p < 0.00025$ (the most conservative adjusted alpha value, 0.05 / [2 + 20 + 180])

[7] The mean token count per vowel/task is N=637 (std. dev., N=426). Token counts are lowest for BOOT (N=979) and highest for BUT (5548); they are highest for self-recordings (N=1983) and lowest for reading passages (N=408). There does not appear to be a over-inflation of significance related to low token counts.

0.0025 (alpha adjusted to 0.05/20). In other words, for all vowels there is at least one task type contrast that is a significant or near-significant predictor of formant variation. Overall, task type is a stronger predictor of variation in F1 than F2. The speech with the greater number and magnitude of differences (relative to interviews) comes from the reading passages. The number and magnitude of differences between Lab Tasks and interview speech are roughly comparable to that between self-recordings and interview speech, though rarely in the same way for the same variable.

Table 2: Effects of elicitation task from 20 separate vowels, 2 formants), for data combined from all three speakers. Adjusted alpha levels: . = 0.05, * = 0.0025, ** = 0.00025, *** = 0.0001.

F1	**BEET**	**BET**	**BAIT**	**BIT**	**BAT**	**BOT**	**BOUGHT**	**BOAT**	**BOOT**	**BUT**
LabTask	*n.s.*	**	*n.s.*	**	*	*n.s.*	*n.s.*	.	*n.s.*	.
Reading	***	***	***	.	***	***	***	***	*	**
SelfRec	***	.	***	.	*n.s.*	.	*	***	**	.
F2	**BEET**	**BET**	**BAIT**	**BIT**	**BAT**	**BOT**	**BOUGHT**	**BOAT**	**BOOT**	**BUT**
LabTask	.	*n.s.*	*n.s.*	*	*	*n.s.*	.	*	*n.s.*	*
Reading	***	.	**	*	***	*	*n.s.*	*n.s.*	*n.s.*	***
SelfRec	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*	*n.s.*	.	*n.s.*	*n.s.*	*n.s.*

One of our main questions asks if Lab Tasks might be a useful replacement for reading passages in field contexts to elicit a speaking style comparable to interview speech while preserving researcher control in eliciting particular lexical items. Here we see some suggestive evidence to support this recommendation: speech from Lab Tasks is generally more similar to speech from interviews than is speech from reading passages. This is a step in the right direction. That said, we still see seven significant (and four marginal) models of difference between Lab Task and interview speech. The remaining question is if those differences might be socially meaningful, which would imply that Lab Tasks and interviews are not sociolinguistically comparable after all. To answer this we must consider speaker-specific models.

4.2 Individual Speaker Models

Data are much more sparse at the individual speaker level, reducing statistical power. Further, inclusion of a random effect of specific Lab Task or specific Reading Passage maximizes the conservatism of these models. Out of 180 models built for this third set of results, only 23 showed a significant effect of elicitation task.[8] For all speakers, differences between interviews and Lab Tasks were much less than the other style contrasts, and only one model showed a

[8] Significant at $p < 0.0025$. An additional 16 models showed a significant effect of elicitation task at $p < 0.05$.

significant difference. These individual-level results suggest that Lab Tasks generally elicit the same speech style as sociolinguistic interviews, whereas self-recorded speech presents a significant stylistic contrast.

4.2.1 Interview speech vs. read-speech

Table 3: All by-speaker models indicating a significant difference between interview speech and reading passage speech at $p < 0.0025$, with a random effect of specific reading passage. All values show a positive effect (larger formant value) for read-speech.

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	F1	F2	F2	F2	--	--	--	--	--
Piper	--	--	F1	--	--	--	--	F1	F1	--
Virginia	F1	--	F1	--	--	--	--	--	--	F1, F2

The results in Table 3 show fewer vowel contrasts between interview speech and read-speech than one might expect from Table 2. Kat and Virginia both show only four of 20 possible effects of task, and Piper only three.[9] This is especially odd for BOT F1 and BOUGHT F1, which were both significantly larger (more open vowels) in read-speech when all three speakers' data was combined (Table 2). Across the differences that do obtain there is a fairly even spread between F1 and F2 effects, although Piper and Virginia favor F1 contrasts while Kat favors F2 contrasts. In all cases an F1 effect indicates a more open vowel in read-speech than interview speech, and an F2 effect indicates a fronter vowel in read-speech than interview speech. The differences possibly suggest the use of a 'clear speech' style in the read tasks (see Smiljanić and Bradlow 2009), resulting in a more expanded vowel space. Kat's BAT F2 effect does not indicate a more expanded vowel space, but may result from a style shift away from the backed variant, which is associated with a stigmatized Californian persona. We return to this idea in §4.2.2.

4.2.2 Interview speech vs. Lab Task speech

Table 4: All by-speaker models indicating a significant difference between interview speech and Lab Task speech at $p < 0.0025$, with a random effect of specific Lab Task. All values show a positive effect (larger formant value) for read-speech.

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	--	--	--	F2	--	--	--	--	--
Piper	--	--	--	--	--	--	--	--	--	--

[9] Seven additional models showed an effect of read versus interview speech at alpha level 0.05: Kat F1 BAT and f2 BEET, BET, and BOT; Piper f1 BET and BOUGHT; Virginia f1 BIT.

Virginia	--	--	--	--	--	--	--	--	--	--

The most striking result in Table 4 is that, despite overall differences in the combined data (Table 2), 59 of 60 individual models showed no significant effect of the task difference between interview and Lab Tasks.[10] The one exception was Kat's production of the BAT vowel in F2 space.[11] The similarity between interview and Lab Task speech may be due to the relaxed interview context. Lab Tasks were treated by both participants as either fun games (Map Task, Diapix) or an extension of the interview; describing a silent film and chuckling about its content is not too dissimilar to describing a memory and chuckling about *its* content. We intentionally sought to control for all contextual factors in order to only compare the difference in task, and it appears that there may be little about the tasks *themselves* that elicits style-shifting. If so, Lab Tasks may prove useful to field sociolinguists who are aiming for a interview style while also eliciting particular lexical items.

Although only one difference obtains for these individual models, it is highly significant, and one that may have a stylistic explanation. Kat (Table 4) produces a fronter production of the BAT vowel in the Lab Tasks than in the interview. Kat is from the San Francisco Bay Area, where BAT backing has been a documented change in progress (Cardoso et al. 2016; Hall-Lew et al. 2015). The backed production has been seen to index a range of social meanings (D'Onofrio 2015, 2018), but particularly meanings associated with the stigmatized Californian persona, 'Valley Girl'. Kat, as a young, female Californian, may vary her production of BAT depending on her motivation to align or disalign herself from this persona. This would explain why Kat produces backed BAT in her interview task, as expected, but fronts the vowel in both read-speech and the Lab Tasks. Anecdotally, Kat explicitly orients differently to the Lab Tasks than the interview, saying: "I know it's not the point...but I wanna win!" This competitive drive may characterize a shift in register demand (Silverstein 2003), invoking a style shift that results in the reduction of a feature that indexes a 'Valley Girl' persona or its related meanings. And while none of the other variables involved in the California Vowel Shift show this same effect for Kat, the speaker analyzed by Authors (2015), who is demographically similar to Kat, showed fronter BIT and BOAT vowels in Lab Task speech than interview speech (and no variation in F1). In terms of the CVS, a fronter BIT vowel is less advanced but a fronter BOAT vowel more advanced.

While results indicate that Lab Tasks can elicit a style comparable to interview speech, there may be exceptions. Descriptively, an analysis of Kat's vowel production based only on speech from Lab Tasks would represent her as a less advanced speaker of this particular regional sound change than would an analysis based on her interview speech, even when the Lab Tasks were

[10] This result contrasts with our preliminary results as presented at NWAV47 (Authors 2018b). Those results defined the phonological context structure differently and did not model a random effect of specific Lab Task.
[11] Even at $p < 0.05$, only three additional variables show a significant difference between interview and Lab Task speech: Kat BOOT F1, Piper BOAT F2, and Virginia BOT F1.

performed by familiar interviewer and in her own home. Lab Tasks might be a useful addition, but not replacement, for interviews.

4.2.3 Interview speech vs. Self-Recorded speech

Table 5: All by-speaker models indicating a significant difference between interview speech and self-recorded speech at $p < 0.0025$, with a random effect of specific self-recording. Negative values and italics indicate a smaller formant value for self-recorded speech.

	BEET	BET	BAIT	BIT	BAT	BOT	BOUGHT	BOAT	BOOT	BUT
Kat	--	--	--	*-F1*	--	--	--	--	--	*-F1*
Piper	--	--	F1	F1	--	F1	--	F1	--	F1
Virginia	--	--	--	--	--	--	--	--	--	F1

In methodological contrast to Lab Tasks are self-recordings. Table 5 shows relatively few differences overall: eight of 60 model comparisons showed a significant difference between interview and self-recorded speech.[12] All effects pertain to variation in F1, not F2.[13] A qualitative comparison between recordings shows that the effects are not due to variation towards different interlocutors in different self-recordings; in all cases the effect of task appears greater than the difference motivated by any particular interlocutor.

Self-recordings differ little from interviews for Virginia, but differ for half of the vowels for Piper. The second author (Virginia's friend) was present at two of Virginia's self-recordings, and the first author (Kat's friend) was present at one of Kat's self-recordings, so the speakers might have been more oriented towards the research motivation of the recording than Piper was in her self-recordings. However, Authors (2017) find considerable variation in sibilant production for all three speakers between self-recordings and interviews, even in those self-recordings with the researchers present.

Kat shows more closed vowels in self-recordings, whereas Piper and Virginia show more open vowels. Kat also shows a higher (fronter) /s/ center of gravity (CoG) in self-recordings, while Piper and Virginia showed lower (backer) CoG values in self-recordings (Authors 2017). It seems that Kat might be orienting towards self-recordings differently than the other two. However, there is no obvious indexical reason why the variables in Table 5 should be more or less susceptible to shifting in self-recordings. In contrast, the speaker in Authors (2015) showed

[12] Six additional models were significant at $p < 0.05$: Kat F1 BOAT and F2 BEET; Piper F1 BOOT and BET; Virginia F1 BAIT and BEET.

[13] An anonymous reviewer notes that this could be due to the comparisons being based on normalized data and/or overall changes in f0 across contexts. While the former should affect all style comparisons (i.e., 4.2.1-4.2.3) in the same way, it is reasonable to posit that f0 variability varies more for self-recordings than all other contexts. Unfortunately, f0 data was not collected.

more advanced CVS front vowels in self-recordings; she especially showed a more open BIT, whereas Kat shows a more closed BIT vowel.

The results for vocalic variation in self-recordings versus interviews are inconclusive, with large individual differences in the amount and direction of difference. In Authors (et al. 2015; 2017) we argued that self-recordings can provide empirical evidence for a wider range of phonetic production than is otherwise obtained from other methods. The results here do not negate that argument, but do highlight Meyerhoff et al.'s (2015) point that interpreting self-recording effects can be particularly challenging.

5 Conclusion

This paper has considered the role of speech elicitation method with respect to phonetic and sociolinguistic methodology. In comparing vowel variation for three speakers across four different elicitation techniques, we find relatively few differences at the individual level, though the differences that do obtain are potentially meaningful. Although we should be careful to generalize from this limited dataset, the results point to the potential benefits of expanding the range of typical speech elicitation methods used in sociophonetic research.

While Lab Task speech and interview speech are differentiated for a third of the vocalic variables, when data are combined, we found only one such difference in individual speaker models. This indicates that the register demand of a Lab Task may not be phonetically indexed differently from a sociolinguistic interview, suggesting that phoneticians and lab phonologists have devised effective tasks to elicit speech more generalizable to "naturally occurring speech" than read-speech (Baker and Hazan 2011: 761). Differences between Lab Tasks in sociolinguistic fieldwork and Lab Tasks in a more typical laboratory context still remain to be seen. For sociophoneticians, the similarities between interview and Lab Task speech allow for speech data to be gathered across a "conversational speech style" spectrum, with Lab Tasks aiding in the collection of infrequently occurring or highly constrained variables while maintaining a speaking style roughly comparable to interview speech. However, where difference does arise, the indexical meaning of the variant may matter. In this regard, reliance *solely* on Lab Tasks as representative of "natural" speech production patterns risks potentially misrepresenting a speaker's production patterns for variables which are most indexically rich. We therefore suggest implementing Lab Tasks as an additional aspect of sociolinguistic methodology, as opposed to viewing these as viable replacements for sociolinguistic interviews.

Despite the results from self-recorded speech complicating conclusions from our previous work (Authors et al. 2015; Authors 2017), we suggest that self-recordings are still worth the methodological and analytical challenges they bring, because they have such a large potential to enrich both the accuracy of our phonetic descriptions and the depth of our indexical analyses, as

evidenced by previous studies (Podesva 2007; Tseng 2014; Van Hofwegen 2016). We now have data on significant variation between sociolinguistic interviews and self-recordings (Sharma 2011; Authors et al. 2015; Authors 2017), and we now need more research to better theorize self-recordings from the perspective of style and social meaning.

Wagner et al. suggest that "if our methods limit the way data is collected (i.e., the way speech is produced) the explanatory power of the results is also limited" (2015: 4). Lab Tasks and self-recordings provide us with two separate, but equally promising, opportunities for data collection in sociophonetics.

References

Author. 2018a.
Authors. 2018b.
Authors. 2017.
Authors et al. 2015
Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson and Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech*. 34(4): 351-366.
Baker, Rachel, and Valerie Hazan. 2011. DiapixUK: A task for the elicitation of spontaneous speech dialogs. *Behavior Research Methods*, 43(4), 761-770.
Bamberg, Michael, and Virginia Marchman. 1991. Binding and unfolding: Towards the linguistic construction of narrative discourse. *Discourse Processes*, *14*(3), 277-305.
Becker, Kara. 2014. The social motivations of reversal: Raised bought in New York City English. *Language in Society* 43(4): 395-420.
Bell, Allan. 1984. Language style as audience design. *Language in Society*, 13, 145-204.
Brown, Gillian, Anne H. Anderson, Richard Shillcock and George Yule. 1984. *Teaching Talk: Strategies for production and assessment.* Cambridge: Cambridge University Press.
Cardoso, Amanda, Lauren Hall-Lew, Yova Kemenchedjieva, and Ruaridh Purse. 2016. Between California and the Pacific Northwest: The Front Lax Vowels in San Francisco English. In Valerie Fridland, Betsy Evans, Tyler Kendall, and Alicia Wassink, eds. *Speech in the Western States, Volume 1: The Coastal States*, pp. 33-54. Publication of the American Dialect Society. Durham, NC: Duke University Press.
Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
D'Onofrio, Annette. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, *19*(2), 241–256.
D'Onofrio, Annette. 2018. Controlled and automatic perceptions of a sociolinguistic marker. *Language Variation and Change*, 30(2), 261-285.

Deterding, David. 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), 187–196.

Drager, Katie. 2018. *Experimental Research Methods in Sociolinguistics*. Bloomsbury Publishing.

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

ELAN (Version 5.0.0-beta). 2017. Nijmegen: Max Planck Institute for Psycholinguistics.

Fairbanks, Grant. 1966. *Experimental phonetics: selected articles*. University of Illinois Press.

Gordon, Matthew, Paul Barthmaier and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141-174.

Grieser, Jessica A. 2019. Investigating Topic-Based Style Shifting in the Classic Sociolinguistic Interview. *American Speech: A Quarterly of Linguistic Usage* 94(1): 54-71.

Hall-Lew, Lauren and Amanda Cardoso, Yova Kemenchedjieva, Kieran Wilson, Ruaridh Purse, and Julie Saigusa. 2015. San Francisco English and the California Vowel Shift. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 591.

Labov, William. 1966. *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington DC.

Labov, William. 1972. Some Principles of Linguistic Methodology. *Language in Society*, *1*(1), 97–120.

Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89(1): 30-65.

Levon, Erez. 2013. Ethnography and recording interaction. In *Research Methods in Linguistics*, Robert J. Podesva and Devyani Sharma, eds. 195–215. Cambridge: Cambridge University Press.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America.* 49(2B):606–608.

Mallinson, Christine, Becky Childs, and Gerard Van Herk. 2013. *Data collection in sociolinguistics: Methods and applications*. Routledge.

Maniwa, Kazumi, Allard Jongman, and Travis Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America* 125:3962–3973.

Meyerhoff, Miriam, Erik Schleef, and Lauren MacKenzie. 2015. *Doing sociolinguistics: A practical guide to data collection and analysis*. Routledge.

Podesva, Robert J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11:478–504.

Podesva, Robert J. 2011a. Salience and the Social Meaning of Declarative Contours: Three Case Studies of Gay Professionals. J*ournal of English Linguistics* 39:233–264.

Podesva, Robert J. 2011b. The California Vowel Shift and Gay Identity. *American Speech* 86:32–51.

Ravindranath, Maya. 2008. The effect of language shift on a sound change in progress. *University of Pennsylvania Working Papers in Linguistics* 14(2): 17.

Rickford, John R. and Faye McNair-Knox. 1994. Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. *Sociolinguistic perspectives on register*, pp.235-276.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2015. *FAVE (Forced Alignment and Vowel Extraction) 1.1.3*. ZENODO. doi:10.5281/zenodo.9846.

Saisuwan, Pavadee. 2016. *Male Femininity in Thai Among Men who Identify with Non-Normative Male Roles.* Doctoral dissertation, Queen Mary University of London.

Scarborough, Rebecca, Jason Brenier, Yuan Zhao, Lauren Hall-Lew, and Olga Dmitrieva. 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.

Schilling-Estes, Natalie. 2002. Investigating stylistic variation. In J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 375-401). Malden: Blackwell.

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

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23(3-4), 193-229.

Smiljanić, Rajka, and Ann R. Bradlow. 2009. Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass* 3(1): 236–264.

Stanford, James. 2010. The role of marriage in linguistic contact and variation: Two Hmong dialects in Texas. *Journal of Sociolinguistics* 14(1): 89–115.

Tseng, Amelia. 2015. Vowel Variation, Style, and Identity Construction in the English of Latinos in Washington, DC. Doctoral dissertation, Georgetown University.

Tucker, Benjamin, Viktor Kharlamov, and Daniel Brenner. 2016. What's the Zed? The Acoustics of Conversational Fricatives in Mid-Western American English. Paper presented at NorthWest Phonetics and Phonology Conference 2016. University of Oregon, Eugene, Oregon.

Tuomainen, Outi, and Valerie Hazan. 2018. Investigating Clear Speech Adaptations in Spontaneous Speech Produced in Communicative Settings. In: *Challenges in Analysis and Processing of Spontaneous Speech,* pp. 9-25. Research Institute for Linguistics of the Hungarian Academy of Sciences. Nyelvtudományi Intézet, Budapest.

Van Hofwegen, Janneke. 2016. A day in the life: What self-recordings reveal about "everyday" language. Paper presented at NWAV 45, Vancouver, BC.

Varon, Sara. 2007. *Robot Dreams*. New York: First Second.

Wagner, Petra, Jürgen Trouvain, and Frank Zimmerer. 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics* 48: 1-12.

Weinberger, Steven H., and Stephen A. Kunath. 2011. The Speech Accent Archive: Towards a typology of accents. *Language and Computers-Studies in Practical Linguistics*, *73*(1), 265-281.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, 1556–1559.

Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336.