Longitudinal phonetic variation in a closed system

Max Bane, University of Chicago
Peter Graff, MIT
Morgan Sonderegger, University of Chicago

1 Introduction

We present a progress report on a large-scale corpus study of the trajectories of different phonetic parameters over the course of three months in a diverse group of speakers situated in a linguistically and socially closed system. The data analyzed so far, and presented in this paper, consist of 806 Voice Onset Time (VOT) measurements from four contestants in Season 9 of the reality-television show Big Brother (Channel 4, United Kingdom). This show offers a unique opportunity to study "medium term" phonetic change in individuals: longer than studies of accommodation in the laboratory or in conversation (hours to days), but shorter than real-time sociolinguistic studies (years). We show that VOT measurements in this dataset, which consists of conversational speech over the course of three months, show several effects consistent with previous work (e.g. the dependence of VOT on place of articulation). We further show that change in the mean VOT of different speakers is non-linear over time, with speakers' VOT trajectories neither fluctuating around a single mean, nor drifting in a single direction throughout the observed period; in addition, a significant portion of this time dependence can be explained in terms of a perturbation to the social organization of the community of speakers, rather than time *per se*. While this last finding is *post hoc* in the sense that the hypothesis was formalized only after building other models of the data, and is therefore subject to verification with a larger dataset, it presents preliminary evidence that social factors can drive phonetic change in individuals over extended periods of time.

2 Background

2.1 The time-course of phonetic variation

There has been significant interest in the extent to which phonetic variables, such as VOT, change over time in individuals. Previous work on longitudinal phonetic variation can be divided into "short-term" and "long-term" studies.

In short-term laboratory studies using an "imitation" or "shadowing" paradigm (reviewed by Nielsen, 2008; Babel, 2009), a subject's productions are compared before and after exposure to a stimulus consisting of speech that has been somehow

* We would like to thank Paul Brazeau, Edward Flemming, Matt Goldrick, Roger Levy, and Jason Riggle for advice; Alan Yu and James Kirby for comments on earlier drafts; and audiences at CLS 46, MIT LingLunch, and Northwestern University for feedback. We are especially grateful to Channel 4/Endemol for permission to obtain footage.

manipulated; comparison is either via analysis of a phonetic variable (e.g. VOT), a broader acoustic variable (e.g. intensity), or perceptual evaluation. Short-term studies which consider phonetic variables generally show that they are influenced by the exposure, but that the direction and magnitude of the effect are influenced by social and linguistic factors. For VOT, Shockley *et al.* (2004) found that subjects' VOTs for voiceless stops increased following exposure to long-VOT voiceless stops. Nielsen (2008) replicated this effect, but found that VOT did not change following exposure to comparable short-VOT stops; Nielsen hypothesizes subjects did not shorten VOT because this would endanger the contrast with voiced stops. Studies on "convergence" or "accommodation" (reviewed by Shepard *et al.*, 2001; Babel, 2009) have examined whether the speech of conversation partners becomes somehow more similar over the course of interaction. Relatively few such studies have focused on phonetic variables. The general finding is again that convergence occurs, but modulated by social and linguistic factors; Heller and Pardo (2010) show this for VOT.

Long-term studies generally examine a particular phonetic variable, for a small set of individuals, at several points in time. The null hypothesis is that phonetic variables remain essentially stable after the critical period; this is implied by the "apparent time" hypothesis of much sociolinguistic work, which states that "linguistic differences among different generations of a population... mirror actual diachronic developments in the language" (Bailey, 2002). In contrast, a well-known set of studies by Harrington *et al.* (2000, *et seq*) shows that Queen Elizabeth II's vowel space has shifted since the 1960s in ways consistent with changes in Received Pronunciation over this period, and some other studies (e.g. Sancier and Fowler, 1997; Evans and Iverson, 2007) have found changes in phonetic variables for individuals who moved between dialect or language regions. However, it is still not clear how typical such long-term change is, or how large the effects are; a review of real-time sociolinguistic studies over the past 20 years concludes only that "... phonology, even though stable in most of its features across individual life spans, is nonetheless available to some speakers for some amount of modification" (Sankoff, 2005).

Short-term studies are able to pinpoint the cause of any change observed in a phonetic variable, but cannot say how the variable behaves over a longer timespan. Long-term studies determine whether a phonetic variable has changed over a period of time, but not what the variable's behavior was between endpoints. The Big Brother corpus offers an opportunity to study change in the "medium term." The large amount of data available allows us to infer relatively detailed trajectories of individuals' phonetic variables over time. It is not clear from previous work what time-dependence these trajectories should exhibit. A null hypothesis is that each individual's trajectory for a given phonetic variable will randomly vary around an unchanging grand mean. If this is not the case, we can ask what the trajectories look like: linear vs. non-linear, monotonic vs. bi-directional, and so on. In this paper, we track VOT in four speakers over three months, and evaluate several hypotheses

about how their VOT values change over time.

2.2 Social factors

It is well-known that speech is socially stratified, in several senses. Both "macro" (e.g. gender, age; Labov, 2000), and "micro" (e.g. jocks vs. burnouts; Eckert, 2000) social categories, as well as the particular social identity a speaker wishes to project (Podesva, 2006), strongly affect phonetic variables. What is less clear is the mechanism by which phonetic variables come to have social meaning in a speech community.

A primary motivation of many short-term studies of the type discussed above is that socially-conditioned convergence of phonetic variables in conversation can help explain dialect formation and the social stratification of speech (Pardo, 2006); that is, that a socially-conditioned shift in a phonetic variable, repeated over the course of many conversations between members of a speech community, can lead to community-level change (Delvaux and Soquet, 2007). However, a link between laboratory studies and community-level change is needed to show that convergence is a possible source of change. Given the number and diversity of social interactions which take place in a community, it is not clear how to test the hypothesis that a particular (set of) convergence effect(s) has led to a given pattern of stratification, without monitoring the majority of interactions and somehow isolating the community from other speakers.

The Big Brother house offers an opportunity to test for effects of social interaction on phonetic variables in a controlled setting. Because the house is a socially-closed system and there is good data on interaction between housemates, it is in principle possible to explicitly test whether change observed in housemates' trajectories for a phonetic variable can be related to social variables. We leave this direction largely for future work with a larger corpus, but do present preliminary evidence that part of the change observed in the trajectories of the four speakers considered here can be explained in terms of a social perturbation in the house.

2.3 Linguistic covariates of VOT

VOT is one of the best-studied phonetic variables. Many studies, almost exclusively laboratory-based, have quantified how a range of linguistic, temporal, and social variables affect VOT, both within and between speakers (see Auzou *et al.*, 2000; Yao, 2009, for partial reviews). In this paper, we consider four known linguistic covariates of VOT: two which impressionistically seem to be the most frequently-discussed in the literature (place of articulation of the stop, speech rate), and two others (whether the stop occurs in a cluster, frequency of the host word) which could be easily calculated for our data. In this subsection we briefly describe and motivate these covariates; we focus on their effect on voiceless stops, as the VOT data considered below come from voiceless stops only.

First, it is well-known that VOT varies by place of articulation (POA). Velar stops (/k/) tend to have longer VOT than alveolar stops (/t/), which tend to have longer VOT than bilabial stops (/p/). The ordering /p/<t/<k/ holds cross-linguistically, and is thought to be linked to articulatory factors (Cho and Ladefoged, 1999). Although not all studies have found significant differences between all 3 places of articulation, crucially no study has reported a statistically significant partial order contradicting /p/<t/<k/. In particular, previous studies on VOT for English voiceless stops produced by adults agree that /p/<t/, but disagree on whether /t/=k/ or /t/<k/ (Whalen *et al.*, 2007).

Laboratory studies have addressed how VOT depends on speaking rate and whether the following segment is a vowel. VOT decreases with increased speaking rate (e.g. Miller *et al.*, 1986; Volaitis and Miller, 1992), and is longer in stop-liquid clusters (e.g. /kl/, /pr/) than in CV sequences (e.g. /ka/, /pa/), controlling for prosody (e.g. Klatt, 1975).

Finally, it has been reported that words with higher token frequency have shorter mean VOT (for voiceless word-initial stops from the Buckeye corpus; Yao, 2009). This accords with the long-standing finding that frequent words have shorter durations, and are generally more affected by various phonetic reduction processes (see Bell *et al.*, 2009 for a review).

The data presented below consists of VOT measurements from four speakers. We first normalize VOT by speaking rate, then assess whether the resulting normalized measurements are sensitive to POA, following vowel vs. consonant, and frequency in directions consistent with previous work.

We note than VOT has been found to covary with many additional linguistic factors, such as the stress (Lisker and Abramson, 1967) and duration (Port and Rotunno, 1979) of the following vowel. Future work will assess the effects of these parameters on VOT. In the models reported below, we assume that at least part of the effect of additional linguistic variables on VOT is taken into account by modeling word specific variation using a random intercept.

Furthermore, social (e.g. gender; Swartz, 1992; Ryalls *et al.*, 1997) and physiological (e.g. hormonal cycles; Wadnerkar *et al.*, 2006) have been found to influence VOT. We do not consider such speaker-dependent variables because the small number of speakers analyzed here makes it impossible to differentiate between the effects of such variables and differences between individual speakers. Future work will consider data from more speakers, making the inclusion of speaker-dependent variables possible.

3 Big Brother

The data were obtained from broadcasts of the reality television show Big Brother (Season 9, Channel 4, United Kingdom). In this show, a group of 15 contestants (known as "housemates") live together in a single house, under constant video and

audio surveillance, for three months. Every week, housemates nominate each other for "eviction." At the end of the week, the viewing public decides via call-in vote which of the two housemates who have received the most nominations are evicted. The final remaining housemate is declared winner and receives a £100,000.00 cash prize. Channel 4's broadcasting of the show includes an hour-long "daily highlights show," as well as a 24/7 live feed from the cameras placed throughout the house.[1]

Housemates are completely isolated from the outside world and without access to any form of media. The show thus constitutes an ideal natural experiment for the study of linguistic behavior over time. Housemates are in a linguistically and socially closed system over a relatively long period, and it is possible to estimate the linguistic input and output of individual speakers, as well as the social dynamics of the community.

Each housemate continually wears a wireless microphone, and is recorded almost continuously by cameras and microphones planted throughout the house. To maintain a constant recording environment and social context, we limit our analysis to audio taken from "diary room clips." These are short segments of the daily highlight shows, in which housemates go into the diary room alone to discuss events in the house with Big Brother (one of several people throughout the season), who communicates through speakers in the wall. Normally, Big Brother is either silent or asks short questions to elicit longer responses from housemates, resulting in a semi-conversational speech style. Audio quality in the diary room is very good. Each housemate visiting it sits in the same chair, providing a (roughly) constant recording environment. We consider only clips where a housemate enters the diary room alone, as is usually the case; this controls for social context. Additionally, housemates visit the diary room frequently enough (every 1–3 days) to infer detailed time trajectories of phonetic variation.

Each season of Big Brother includes a "twist" which creates unusual social situations for the housemates. The twist in the season considered here was the "Heaven-Hell divide." During Episodes 37–67, the house was split into "Heaven" and "Hell." Each week during this period, each housemate was assigned to Heaven or Hell, depending on performance on various tasks. Hell housemates were compelled to cook for Heaven housemates, and restricted to less comfortable household amenities and curtailed smoking privileges. In contrast, heaven housemates had unrestricted smoking rights and sole access to the "luxury" bathroom and bedroom. This social division was physically embodied by a one-meter-high Plexiglass barrier, preventing Hell housemates from crossing to the Heaven side (and vice versa). While housemates' mobility was constrained by the divide, physical and linguistic interaction across the Plexiglass fence was possible, and occurred frequently. Below, we find that the divide is a social perturbation which coincides with changes in the VOT trajectories of the speakers included in this study.

[1]For brevity, we omit some details of the show's logistics which are not relevant here; for example, two housemates were ejected from the house for bad behavior, rather than voted off.

HM	**Episode** (number of VOT measurements)	Total VOTs
Lisa	**2** (10), **13** (30), **20** (15), **27** (18), **35** (27), **41** (17), **59** (24), **61** (10), **68** (20), **78** (13), **82** (10), **87** (11)	205
Michael	**3** (25), **13** (23), **17** (12), **22** (12), **41** (31), **45** (18), **68** (9), **76** (21), **84** (43), **92** (23)	217
Rachel	**2** (4), **6** (23), **12** (12), **19** (5), **20** (17), **22** (5), **33** (11), **37** (5), **54** (22), **61** (19), **65** (4), **73** (6), **76** (11), **87** (2), **90** (21), **92** (20)	187
Rex	**8** (26), **15** (35), **27** (8), **41** (30), **51** (18), **59** (17), **76** (25), **84** (14), **92** (24)	197

Table 1: Distribution of VOT tokens measured across speakers and episodes.

4 Data

Due to weekly evictions, the number of speakers in the house decreases over time. The data used here are from four housemates present for most of the three-month season: Rachel (the winner), Michael, Lisa, and Rex. Of these four, Lisa is evicted in the 87^{th} episode, while the rest are present until the 92^{rd} and final episode. The data are VOT measurements on word-initial, voiceless plosives (/p/, /t/, /k/) extracted from diary room clips (described above). The temporal distribution of these measurements is summarized in Table 1. The overall number of VOT measurements for each housemate is roughly balanced, at 200 ± 15 measurements.

VOT was defined as the duration from the burst of the plosive release to the zero-amplitude crossing of the first unambiguous period of subsequent voicing. Measurements were made by two of the authors (MB, MS). A subset of the tokens (13%) was measured by both, in order to estimate the degree of inter-transcriber agreement; the two transcribers' measurements had correlation $r = 0.91$, and mean absolute difference of 4.5 ms.

To control for the significant effect of speaking rate on VOT (see Section 2.3), VOT was normalized (i.e. multiplied) by the number of syllables produced per second in a "spurt" of connected speech.[2] A spurt was defined as a contiguous segment of speech containing no interval of silence longer than 0.4 seconds. Within a spurt, syllables were counted automatically using the algorithm described by de Jong and Wempe (2009), which searches for local peaks of intensity as potential syllable nuclei, discounting those that lack sufficient periodicity or a minimum pitch, in order to exclude high-intensity consonants. In the terminology of de Jong and Wempe, we required a minimum pitch of 50 Hz, an "ignorance level" of 0 dB (the amount by which an intensity peak must exceed the median intensity of the spurt), and a mini-

[2]Thus the interpretation of our normalized VOT measurements is milliseconds $\times$ (syllables/milliseconds), i.e. a fraction of the average syllable in the surrounding spurt.

mum dip between intensity peaks of 2 dB. The algorithm was found by de Jong and Wempe to correlate well with hand-measured syllable counts in corpora of Dutch speech ($0.7 \leq r \leq 0.88$)

We note that previous laboratory studies have used a variety of measures of speech rate, such as following vowel duration and surrounding syllable duration. However, these measures could in principle vary independently, and it is not yet certain how they extend to conversational speech, making it unclear which measure(s) to use for our corpus. We use only one measure of speech rate here, but plan to consider several in a more principled fashion in future work.

5 Models

In this section, we employ linear mixed effects modeling to test different hypotheses about the VOT distributions of our four speakers over the course of the three-month season. We test (i) whether VOT measurements obtained so far exhibit the same effects of different linguistic factors found in previous studies, and (ii) how (if at all) housemates' VOT distributions change over time. The dependent measure in all models is VOT normalized by speech rate, as described above. Unless otherwise noted, we use "VOT" to refer to normalized VOT below.

5.1 Predictors

To test whether VOT measurements in our corpus show effects of the linguistic factors discussed above (Section 2.3) in accord with previous work, we include a number of predictors based on these factors. PLACE OF ARTICULATION of the word-initial consonant (bilabial, alveolar, velar) was Helmert-coded, comparing mean VOT for bilabial (/p/) to alveolar stops (/t/) and mean VOT for [+anterior] (/p/, /t/) to [−anterior] stops (/k/) . One contrast-coded predictor (PHONOLOGICAL CONTEXT) assessed the difference in VOT between stops in CV sequences vs. in stop-liquid clusters (*peas* vs. *please, pray*). Finally, the centered log token frequency (CELEX2; Baayen *et al.* 1996) of the particular word beginning in a voiceless stop was included as a numerical predictor, to test the effect of frequency on VOT (FREQUENCY). To estimate the effect of time on VOT trajectories, we include a numerical predictor: the number of days elapsed in the season at utterance time (TIME); this predictor was centered.

To control for speaker-specific differences in VOT, we further included a dummy-coded predictor specified for the particular speaker (SPEAKER), which was then contrast coded (i.e. each of the 3 SPEAKER contrasts sums to 0 across the dataset). This is preferable to using random effect terms (i.e. by-SPEAKER random intercepts or random slopes), given the small number of levels (4) for SPEAKER in the current dataset (Snijders and Bosker, 1999). We also included (two-way) interactions of SPEAKER with FREQUENCY, PLACE OF ARTICULATION, and TIME, as each of these terms significantly improved data likelihood ($p < 0.05$). The interaction of

Predictor	Coding	d.f.	χ^2	p
SPEAKER	dummy	15	132	<0.0001
+ all interactions				
TIME	numerical	4	11.2	<0.05
+ interactions with SPEAKER				
PLACE OF ARTICULATION	Helmert	8	57.3	<0.0001
+ interactions with SPEAKER				
PHONOLOGICAL CONTEXT	treatment	1	29.6	<0.0001
LOG(FREQUENCY)	numerical, centered	4	31.3	<0.0001
+ interactions with SPEAKER				
WORD (random effect)	none	1	16.0	<0.0001

Table 2: Predictors and their contribution to data likelihood.

SPEAKER with PHONOLOGICAL CONTEXT was not included as it did not significantly improve data likelihood ($\chi^2(3) = 7.1$, $p = .068$). To allow for word-specific variation in VOT, we included a predictor (WORD) indexing the particular word containing the measured consonant, as a random intercept (235 levels: *to*, *peas*, *can*...).

The different predictors included in the model, as well as their contributions to data likelihood, are summarized in Table 2. Overall collinearity of predictors was very low. The average partial correlation of fixed effects was .09 and the highest variance inflation factor was 2.08 (where 4, 7 and 10 are common cut-off points for trustworthy estimates; Fox, 1991). We can therefore be confident that the effect sizes and directions reported in the next subsection are reliable.

5.2 Results

The linear mixed effects model was fitted in `R` using the `lmer()` function, from the `lme4` package for mixed-effects models (Bates and Maechler, 2010). We summarize the model, omitting all interactions for brevity.[3]

We first consider the estimates for the main effects of the linguistic predictors. The effects of PHONOLOGICAL CONTEXT and FREQUENCY were as expected from previous studies. VOT was higher for words with higher token frequency ($\beta = -0.0038$, $t = -1.84$, $p_{MCMC} < 0.01$), and normalized VOT was higher for stops occurring in stop-liquid clusters (mean=0.30, SD=0.11) than for stops occurring in CV sequences (mean=0.24, SD=0.09; $\beta = 0.063$, $t = 5.60$, $p_{MCMC} < 0.0001$). The averages for non-normalized VOT were 77 ms (SD=29) and 60 ms (SD=23) for clusters and CV sequences, respectively. Figure 1 illustrates

[3]The model formula in `lme4`-style is: VOT.NORMALIZED $\sim$ PLACE*SPEAKER + PHON.CONTEXT + log(FREQUENCY)*SPEAKER + TIME*SPEAKER + (1|WORD), where predictors are coded as described in the text.

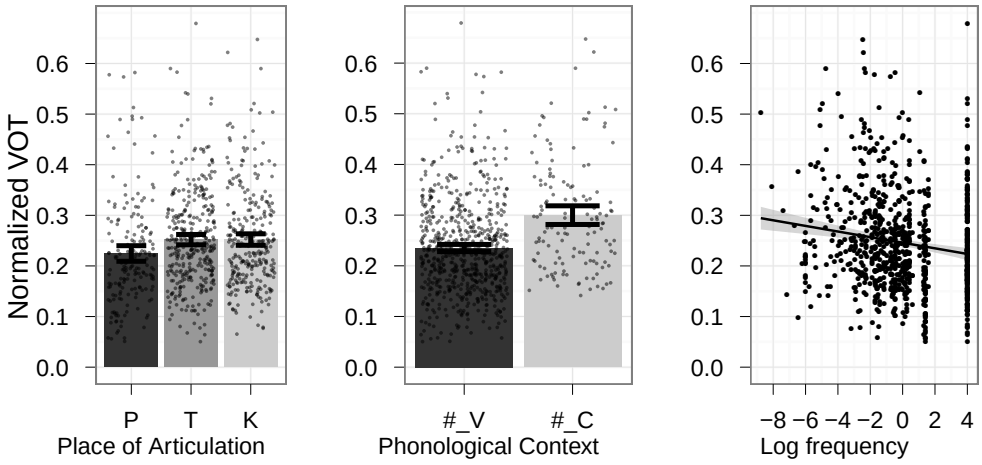

Figure 1: Mean normalized VOT depending on linguistic factors investigated.

the means of normalized VOT in different phonological contexts, and its correlation with word frequency.

While not substantiating the full /p/</t/</k/ dependence of VOT on place of articulation often discussed in the phonetic literature, the model's estimates for PLACE OF ARTICULATION are consistent with previous laboratory studies. VOT is higher for /t/ than for /p/ ($\beta = 0.045$, $t = 3.60$, $p_{MCMC} < 0.0001$), but there is no significant difference in VOT between [-anterior] (/k/) and [+anterior] (/p/, /t/) stops ($\beta = -0.0045$, $t = -0.49$, $p_{MCMC} = 0.20$). Our data thus show the partial order /p/</t/=k, the same result as some previous laboratory studies on VOT in English (see Whalen *et al.*, 2007). Mean normalized VOT was 0.22 (SD=0.10), 0.25 (SD=0.096) and 0.25 (SD=0.098) for /p/, /t/ and /k/ respectively. The non-normalized VOT means are 56 ms (SD=25 ms), 65 ms (SD=24 ms), and 64 ms (SD=25 ms). Normalized VOT means for stops at different places of articulation are illustrated in Figure 1.

Overall, the estimates for the linguistic predictors included in the model are consistent with observations from previous laboratory studies. This consistency is encouraging, given the conversational nature of the speech measured for this study.

Additionally, there is a significant main effect of SPEAKER (Lisa $\beta = -0.0025$, $t = -0.46$, $p_{MCMC} = 0.79$; Michael $\beta = -0.0014$, $t = -0.27$, $p_{MCMC} = 0.51$; Rachel $\beta = 0.031$, $t = 5.6$, $p_{MCMC} < 0.0001$), meaning that some speakers differ in their intrinsic VOT ranges, after the effects of linguistic factors, speech rate, and time have been accounted for; this agrees with laboratory findings by Allen *et al.* (2003). The by-word random intercept significantly contributes to data likelihood ($\chi^2(1) = 16.2, p < 0.0001$), indicating that words in the dataset have different VOT ranges, after the effects of some word specific properties (i.e. place of articulation of the stop) have been accounted for. Including the random intercept reduces the

chance that word-specific differences in VOT or certain outliers in the data drive the significances reported here and at least partially accounts for the effect of linguistic variables not considered in this study.

Finally, we observe that a linear effect of time, including interactions with the different levels of SPEAKER significantly improves data likelihood (see Table 2). Figure 2 (top) depicts LOWESS-smoothed lines of VOT measurements for the four speakers over time as well as the predictions of the model, with its linear predictor for time. (Model predictions are shown with all predictors other than SPEAKER and TIME held at their mean values).

It is immediately apparent that the fit of the linear model-predicted curves to the smoothed trajectories is rather poor. Speakers' VOT values do not appear to be either randomly varying around a single mean over the course of three months, nor changing monotonically (linearly or otherwise). Speakers' trajectories both increase and decrease during different parts of the season. To capture this behavior, a model allowing for more complex time dependence is needed. We now attempt to improve the model's predictions for the effect of time, using model comparison to test whether the addition of more complex time dependence is justified by the data.

5.3 Improving predictions over time

The first step we take to improve the model is to allow for a non-linear effect of time on speakers' VOT trajectories. We do so by including restricted cubic splines (RCS) of TIME, using `rcs()` in the `R` package `Design` (Harrell, 2009). Restricted cubic splines allow a non-linear function of a continuous predictor to be included in a regression model without overfitting to peripheral values of the predictor (as can occur if polynomial terms such as x^2 are included). The non-linear function is parametrized by the number k of "knots" it includes; roughly, $k-2$ is the number of "turns" the function has. RCS with 2 knots is simply a linear term; RCS with 3 knots is analogous to a quadratic function (in that it only turns once), and so on.[4] We denote a restricted cubic spline with k knots as RCS_k.

Of course, a model which allows for more complex functions of a continuous predictor will always achieve better coverage, simply by having more degrees of freedom. The question is whether the added complexity is justified in terms of the improvement in data likelihood which is achieved. The χ^2-likelihood test allows us to assess the improvement in data likelihood relative to the added complexity. We first test a model where TIME is coded as RCS_3, and find that model fit is improved significantly ($\chi^2(4) = 20.4$, $p < 0.001$) relative to a model with only a linear term. However, a model where TIME is coded as RCS_4 does not exhibit a significantly improved fit ($\chi^2 = 1.3$, $p = 0.86$) relative to RCS_3: the addition of a fourth knot is not justified relative to the added complexity.[5]

[4] For more detail see Harrell (2001, 16–24), and Baayen (2008, Sec. 6.2.1).

[5] Strictly speaking, only differences in data likelihood between nested models are distributed as χ^2. RCS_4 consists of 3 basis functions of time, $R_4 = (r_1^4(t), r_2^4(t), r_3^4(t))$ and RCS_3 consists of 2,

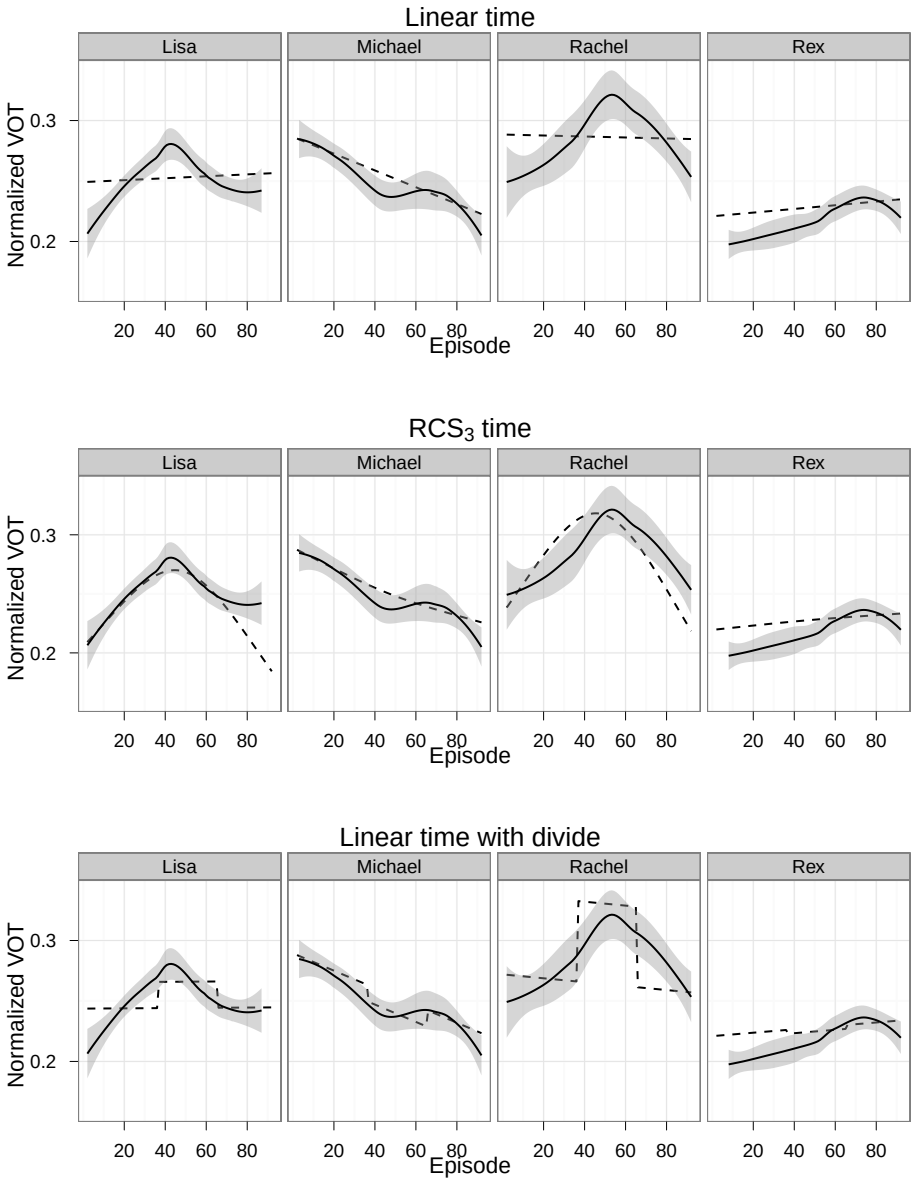

Figure 2: Empirical VOT trajectories vs. model-predicted trajectories for models with time coded linearly (top), as RCS_3 (center), or as a linear term + DIVIDE (bottom). For each model, empirical LOWESS-smoothed VOT trajectories (solid) with standard error (shading) are compared against model-predicted trajectories (dashed). Model predictions are with all predictors other than SPEAKER and TIME held at their mean values.

The resulting model, where TIME is coded as a restricted cubic spline with 3 knots, agrees markedly better than a linear model with the empirical time trends observed in our data. For each speaker, the predictions of the RCS_3 model and the empirical LOWESS-smoothed VOT time trajectories are compared in Figure 2 (middle).

This result constitutes strong evidence for non-linear change in phonetic parameters over time. Speakers' VOT distributions are neither constant in the sense that they fluctuate around a single mean; they show non-linear and non-monotonic (i.e. not unidirectional) time trends. What is not clear at this point is the source of the time trends observed in speakers' VOT trajectories. We next present evidence that the non-linear portion of this time dependence can be accounted for by a social variable: the Heaven-Hell divide (discussed in Section 3).

5.4 Towards an explanation for non-linear trajectories

In what follows we introduce a predictor based on a social variable relevant in the Big Brother house, and use it to perform a *post hoc* revision of the last model (where time was coded non-linearly, as RCS_3). In the revised model, much of the change in housemates VOT trajectories can be explained through this variable, rather than the effect of "time" as such.

Figure 3 shows LOWESS-smoothed normalized VOT trajectories of the four speakers, as well as the beginning and end of the presence of the divide. We observe that some housemates' trajectories (those of Lisa and Rachel, and possibly also Michael's) exhibit an inflection point as well as increased slopes somewhere between Episodes 40 and 60. This corresponds with the period when trajectories are most dispersed from each other, having begun closer together, and finishing the season possibly closer still. We note *post hoc* that this period roughly coincides with a significant social perturbation to the community of speakers: the presence of the Heaven-Hell divide.

Inspired by this co-incidence, we test whether the non-linear time dependence in our model can be accounted for by the presence of the Heaven-Hell divide. To allow this factor to affect VOT trajectories, we use a dummy-coded predictor (DIVIDE) and its interactions with SPEAKER; this allows the model to add or subtract a constant from a given speaker's modeled VOT while the divide is present. DIVIDE is 1 when the divide is present (when $37 \leq$ TIME ≤ 67), and 0 otherwise.

In order to test whether non-linearities over time can be accounted for by a main effect of DIVIDE, we employ non-nested model comparison using the χ^2-likelihood test. We first fit a superset model containing both DIVIDE and time coded as RCS_3 (as well as their interactions with speaker).[6] We remove each set of predictors

$R_3 = (r_1^3(t), r_2^3(t))$. Because r_2^3 is a slightly different function from r_2^4, $R_3 \nsubseteq R_4$. However, for our data the correlation between r_2^3 and r_2^4 is > 0.998, and we consider the models close enough to use nested model comparison.

[6]Superset model formula: VOT.NORMALIZED $\sim$ PLACE *SPEAKER + PHON.CONTEXT +

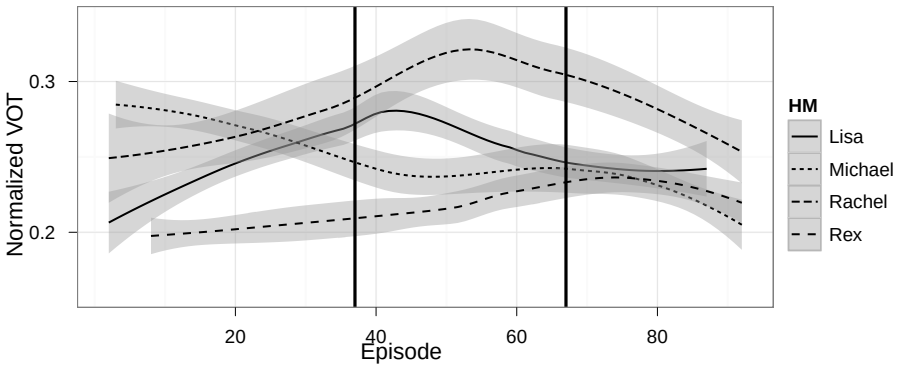

Figure 3: Empirical LOWESS-smoothed VOT trajectories; vertical lines mark the beginning and end of the presence of the Heaven-Hell divide.

(DIVIDE+interactions, RCS_3+interactions) in turn,[7] one at a time, to test whether there is a significant difference in data likelihood between the superset model and each of the subset models. If one of the subset models differs significantly from the superset model, while the other one does not, we can conclude that the predictor set in the model that accounts for the same amount of variance as the superset model is sufficient to account for the variance in the data.

We observe no significant difference in data likelihood between the superset model and the subset model including DIVIDE ($\chi^2(4) = 5.6$, $p = 0.23$), and a marginally significant difference between the superset model and the subset model including the RCS_3 of time ($\chi^2(4) = 9.3$, $p = 0.054$). By the logic outlined above, we can conclude that DIVIDE accounts for most of the variance otherwise accounted for by RCS_3, and not vice-versa. We note that non-nested model comparison using the Bayesian Information Criterion (BIC) gives analogous results.[8]

We have thus shown that a variable related to interaction between housemates, DIVIDE, can account for the non-linear effect of time observed in our data. There are caveats: as divide was coded after observing the peculiar shape of VOT trajectories, this result must be verified against data from more speakers, particularly given that the result of non-nested model comparison is only marginally significant; additionally, we have not shown that a predictor based on the presence of the Heaven-Hell divide is preferable to any *other* linear adjustment of VOT trajectories for some fixed time interval (e.g. for 20≤TIME≤60); this could be tested by fitting a "breakpoint" regression model. Nonetheless, the finding that perturbations to our mini speech community's social system can potentially explain part of the

log(FREQUENCY)*SPEAKER + RCS_3(TIME)*SPEAKER + DIVIDE*SPEAKER + (1|WORD).

[7] When RCS_3+interactions are removed, they are replaced by linear TIME+interactions.

[8] The model with time coded as RCS_3 has BIC=1676. The model including DIVIDE has BIC=1674, where lower values imply better fit given model complexity.

time dependence of a phonetic variable in that community is novel, and preliminary evidence for the significance of our data for the study of community-level phonetic change.

6 Discussion and Conclusion

The three-month timescale of the Big Brother corpus lies somewhere between the short-term (hours to days) and long-term (months to years) timescales considered in previous work on longitudinal phonetic change in individuals. With "medium-term" observational studies of this sort, it is possible to consider relatively frequent measurements over significant time periods. Furthermore, the unique format of the Big Brother reality show allows us to track the phonetic variation of a community that is strictly isolated from outside linguistic contact. We therefore expect the continued collection and analysis of such data to illuminate the structure and causes of language dynamics in the house.

The preliminary dataset analyzed here allows some qualitative conclusions that are statistically robust. First is that the longitudinal distribution of conversational VOT can be modeled as conditional on at least some of the same covariates as are known to affect VOT in previous work (primarily laboratory studies): the individual speaker, his/her speaking rate, the place of articulation, whether the following phone is C or V, and word frequency. Second is that speakers' VOT time trajectories do not appear to be random fluctuation around constant means, at least not as would be evident on the timescale of the study.[9] Furthermore, speakers' trajectories do not change at a constant rate, and are seen to reverse directions at points.

Finally, the data are at least consistent with a *post hoc* sociological hypothesis about changes in speakers' trajectories—that they coincide with the Heaven/Hell division of the house. The obvious question is *why* DIVIDE affects VOT trajectories as it does. We can currently only offer speculations on this front, to be tested in the larger dataset. Housemate trajectories seem to diverge when the divide is present; this could somehow be due to increased conflict in the house during this period. It is intriguing that the effect of DIVIDE on trajectories seems to differ by housemate gender: positive and higher amplitude for women, negative and lower amplitude for men. The small number of speakers in the current dataset currently prevents us from differentiating between a gender effect and random differences between speakers, but the effect of social variables on VOT trajectories is an exciting direction for future work. The effect of the divide could also be related to an aspect of housemate dynamics not considered here, such as accommodation to housemates on the same side of the divide, or changed frequencies of interaction between each pair of housemates during the split. Thanks to the amount of data available on housemate interactions, it should be possible to test each of these hypotheses in future work.

[9] That is, the observed time trajectories could in fact be part of a process of random variation, but on a much longer timescale than 3 months.

Regardless of its cause, the effect of DIVIDE gives some encouragement that a larger dataset will reveal further relationships between social dynamics and phonetic change in the house. In the long term, we believe that the study of phonetic change in a socially and linguistically closed system can enrich our understanding of the time-course of phonetic variation and the effect of social variables on phonetic variables, and potentially inform theories of dialect formation and the origin of social stratification in speech.

References

Allen, J.S., J.L. Miller, and D. DeSteno. 2003. Individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 113.544–552.

Auzou, P., C. Ozsancak, R.J. Morris, M. Jan, F. Eustache, and D. Hannequin. 2000. Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical Linguistics & Phonetics* 14.131–150.

Baayen, R.H. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.

——, R. Piepenbrock, and L. Gulikers. 1996. *CELEX2 (CD-ROM)*. Philadelphia: Linguistic Data Consortium.

Babel, M.E., 2009. *Phonetic and Social Selectivity in Speech Accommodation*. UC Berkeley dissertation.

Bailey, G. 2002. Real and apparent time. In *The Handbook of Language Variation and Change*, ed. by J.K. Chambers, P. Trudgill, and N. Schilling-Estes, 312–331. Malden, MA: Blackwell.

Bates, D., and M. Maechler, 2010. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-34.

Bell, A., J.M. Brenier, M. Gregory, C. Girand, and D. Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60.92–111.

Cho, T., and P. Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *J. Phonetics* 27.207–229.

de Jong, N.H., and T. Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41.385–390.

Delvaux, V., and A. Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64.145–173.

Eckert, P. 2000. *Linguistic variation as social practice*. Malden, MA: Blackwell.

Evans, B.G., and P. Iverson. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *J. Acoust. Soc. Am.* 121.3814–3826.

Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks, CA: Sage.

Harrell, Jr., F. 2001. *Regression modeling strategies*. New York: Springer.

——, 2009. *Design: Design Package*. R package version 2.3-0.

Harrington, J., S. Palethorpe, and C.I. Watson. 2000. Does the Queen speak the Queen's English? *Nature* 408.927–928.

Heller, J., and J. Pardo. 2010. First impression and speaker role influence VOT convergence in conversation. Talk presented at 16th Mid-Continental Workshop on Phonology. Paper in preparation.

Klatt, D.H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *J. Speech Lang. Hear. R.* 18.686–706.

Labov, W. 2000. *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.

Lisker, L., and A.S. Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10.1–28.

Miller, J.L., K.P. Green, and A. Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43.106–115.

Nielsen, K.Y., 2008. *Word-level and Feature-level Effects in Phonetic Imitation*. UCLA dissertation.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119.2382–2393.

Podesva, R., 2006. *Phonetic Detail in Sociolinguistic Variation*. Stanford dissertation.

Port, R.F., and R. Rotunno. 1979. Relation between voice-onset time and vowel duration. *J. Acoust. Soc. Am.* 66.654–662.

Ryalls, J., A. Zipprer, and P. Baldauff. 1997. A preliminary investigation of the effects of gender and race on voice onset time. *J. Speech Lang. Hear. R.* 40.642–645.

Sancier, M.L., and C.A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *J. Phonetics* 25.421–436.

Sankoff, G. 2005. Cross-sectional and longitudinal studies in sociolinguistics. In *Sociolinguistics: An international handbook of the science of language and society*, ed. by U. Ammon, N. Dittmar, K.J. Mattheier, and P. Trudgill, volume 2, 1003–13. Berlin: de Gruyter.

Shepard, C.A., H. Giles, and B.A. Le Poire. 2001. Communication accommodation theory. In *The new handbook of language and social psychology*, ed. by W.P Robinson and H. Giles, 33–56. New York: Wiley.

Shockley, K., L. Sabadini, and C. A. Fowler. 2004. Imitation in shadowing words. *Percept. Psychophys.* 66.422–429.

Snijders, T.A.B., and R.J. Bosker. 1999. *Multilevel analysis*. Thousand Oaks, CA: Sage.

Swartz, B.L. 1992. Gender difference in voice onset time. *Perceptual and motor skills* 75.983–992.

Volaitis, L.E., and J.L. Miller. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.* 92.723–735.

Wadnerkar, M.B., P.E. Cowell, and S.P. Whiteside. 2006. Speech across the menstrual cycle: A replication and extension study. *Neurosci. Lett.* 408.21–24.

Whalen, D.H., A.G. Levitt, and L.M. Goldstein. 2007. VOT in the babbling of French-and English-learning infants. *J. Phonetics* 35.341–352.

Yao, Y. 2009. Understanding VOT variation in spontaneous speech. In *Proc. 18th International Congress of Linguists (CIL XVIII)*, 29–43. Seoul: Korea University.