THE MEDIUM-TERM DYNAMICS OF ACCENTS ON REALITY TELEVISION

Morgan Sonderegger
McGill University and *Centre for Research on Brain, Language, and Music*

Max Bane
University of Chicago

Peter Graff
University of Vienna

How flexible is an individual's accent during adulthood, and how does this flexibility relate to longer-term change? Previous work has found that accents are remarkably flexible in conversational interaction, but predominantly stable over years, leading to very different views of the role of individuals in community-level sound change. This article examines MEDIUM-TERM accent dynamics (days to months) by taking advantage of a 'natural experiment': a reality television show where contestants live in an isolated house for three months and are constantly recorded, forming a closed system where it is possible to both determine the dynamics of contestants' speech from day to day and reason about the sources of any observed changes. We build statistical models to examine time dependence in five phonetic variables within individuals, in 14.5 hours of spontaneous speech from twelve English-speaking contestants. We find that time dependence in pronunciation is ubiquitous over the medium term: large daily fluctuations in pronunciation are the norm, while longer-term change over weeks to months occurs in a minority of cases. These patterns mirror the conflicting findings of previous work and suggest a possible bridge between the two. We argue that time dependence in phonetic variables is influenced by contrast between sounds, as well as systematic differences between speakers in how malleable their accents are over time; however, we find only limited evidence for convergence in individuals' accents. Our results have implications for theories of the role of individuals in sound change, and suggest that medium-term pronunciation dynamics are a fruitful direction for future work.*

Keywords: longitudinal studies, accent dynamics, phonetic variation, individual differences, language change

1. Introduction. Two of the most striking aspects of language are stability and change. Both aspects are especially clear for how sounds are pronounced, the focus of this article. At the level of individuals, we intuitively know that a speaker's accent can shift as a function of her interlocutor, and that some people's accents seem to shift permanently (ACCENT CHANGE), for example, after moving to a new dialect region. However, we also intuitively know that many people's accents are stable over time: this default assumption underlies our ability to identify where someone is from by their accent. At the level of speech communities, many aspects of the phonetic and phonological system are stable at a given point in time, yet cases of SOUND CHANGE are always occurring (e.g. Labov 1994). This article addresses three broad questions about accent dynamics and sound change: how plastic are the accents of individual speakers during adulthood, what are the sources of the dynamics of a speaker's accent, and what is the relationship of accent plasticity in individuals to community-level sound change?

* We thank Molly Babel, Lauren Hall-Lew, Tyler Kendall, Jane Stuart-Smith, Natasha Warner, and an anonymous referee for helpful comments on drafts of this article, and the many other people who have provided useful feedback on this project, including Peter Auer, Rina Foygel Barber, Paul Brazeau, Alex Fine, James Kirby, Matt Goldrick, John Goldsmith, Dan Jurafsky, Robin Queen, Jason Riggle, Tyler Schnoebelen, Charlotte Vaughn, and Alan Yu; as well as audiences at CLS 2010, LSA 2011, CUNY 2011, NWAV 2012, LabPhon 2014, ICPhS 2015, and at USC, Indiana U., U. Michigan, McGill, U. Edinburgh, U. Ottawa, U. Minnesota, U. Toronto, Cornell, Northwestern U., and MIT. We thank N. Rothfels, M. Nelson, T. Knowles, M. Schwarz, M. Labelle, L. Bassford, R. Hwang, H. Hsieh, and M. Coumel for research assistance. We gratefully acknowledge permission from Channel 4/Endemol to access footage from *Big Brother 9 UK*. Preliminary versions of the VOT results were given in Bane et al. 2010 and Sonderegger 2015, and in Sonderegger 2012 for all variables. This work was supported by grants from the Social Sciences and Research Council of Canada (#430-2014-00018) and the Fonds de Recherche du Québec Société et Culture (#183356).

Previous work has addressed these questions on two timescales: SHORT-TERM shifts in pronunciation during conversation and laboratory experiments (seconds–hours), and LONG-TERM accent change over the lifespan (years).

1.1. SHORT-TERM. One long-standing body of work emphasizes the flexibility of the sound systems of individuals, which are constantly updated over time as a result of interaction (Hockett 1965, Paul 1880; see Garrett & Johnson 2013). For example, contemporary EMERGENTIST theories of phonetics and phonology (e.g. Johnson 1997, Pierrehumbert 2001), whose core tenets go back to the Neogrammarians, assume that a speaker's cognitive representation of sounds includes all examples ever heard. This representation is updated with each new interaction and used in speech production. Thus, emergentist models predict that pronunciation norms can and will shift over time, including during adulthood, as a result of interactions.

A primary source of evidence for accent plasticity in adults comes from short-term shifts in how one speaks (over seconds–hours), in conversations and in laboratory settings, under exposure to other people's speech. These shifts (termed imitation, (phonetic) convergence, accommodation, etc.) have been documented using a variety of paradigms (Pardo 2013). As assessed by perceptual measures, subjects shift their overall pronunciations toward a target voice in word-shadowing tasks (Fowler et al. 2003, Goldinger 1998) and toward their interlocutor in conversational interaction (e.g. Pardo 2006). Specific acoustic variables typically also undergo short-term shifts in both interactive and noninteractive settings. For example, American English speakers increased the VOT of word-initial voiceless stops toward a model talker in an imitation task (Nielsen 2011) and shifted their productions of five vowels (in F1/F2 space) toward a model talker in a shadowing task (Babel 2011). Pardo (2009) also found that American English speakers shifted vowel formants in conversational interaction, but in more complex ways. In these and other studies of shifts in phonetic variables, the existence of some shift, as a result of exposure to another person's speech, is quite robust across speakers.

Short-term shifts in pronunciation are to some extent 'automatic' consequences of interaction (Delvaux & Soquet 2007, Goldinger 1998). They are also heavily modulated by three types of factors, which result in the significant variability between speakers in the size and directionality of shifts observed in most short-term studies:

(i) SOCIAL FACTORS, such as attitude toward the interlocutor and gender (Babel 2010, Bourhis & Giles 1977, Namy et al. 2002). COMMUNICATION ACCOMMODATION THEORY (Giles et al. 1991, et seq.) proposes that these shifts result from individuals managing social distance using an ACCOMMODATION strategy, such as CONVERGENCE or DIVERGENCE (e.g. to express solidarity or disapproval).

(ii) LINGUISTIC FACTORS. For example, Nielsen (2011) cites contrast maintenance to explain why English speakers will shift VOT of voiceless stops toward lengthened but not shortened VOT (cf. Babel 2011, Mitterer & Ernestus 2008).

(iii) INDIVIDUAL DIFFERENCES, correlated with factors such as personality traits and gender (Namy et al. 2002, Yu et al. 2013).

In sum, plasticity in pronunciation, modulated by multiple factors, is the norm over seconds–hours for adults. This plasticity is in line with the ubiquity of STYLE SHIFTING: shifts in a speaker's linguistic usage as a function of the addressee, topic, and so forth (e.g. Bell 1984, Eckert 2000, Rickford & McNair-Knox 1994), possibly many times over the course of a day (e.g. Coupland 1980, Hindle 1980).

Change by accommodation. The existence of substantial accent plasticity in interactions goes naturally with the view that change during adulthood is an important driver of sound change in communities. Versions of this view go back to the Neogrammarians (Paul 1880), as summarized by Auer and Hinskens (2005), who call it the change by accommodation (CBA) model. The CBA model explains the relationship between individual-level and community-level change as a sequence of three steps: (1) people accommodate to each other during interactions; (2) eventually, a given individual's norms change as the result of accumulation of these interactions; (3) the innovation spreads in the wider community (as individuals undergo step 2). Different versions of the CBA model address dialect change/contact settings alone, or all sound change (e.g. Bloomfield 1933, Garrett & Johnson 2013, Labov 1990, Trudgill 1986, 2004), and differ in whether short-term shifts (and thus steps 2–3) are fundamentally automatic (every interaction makes the sound system of interlocutors more similar) or fundamentally social (as proposed by communication accommodation theory). Regardless of their mechanism, short-term shifts are generally assumed to be 'the driving force of language change' (Auer & Hinskens 2005:356). The fact that most short-term literature considers adults suggests that steps 1–2 refer by default to adults. Thus, the link being made between accent plasticity in adulthood and the nature of sound change is this: because adults' pronunciations are so plastic in interactions (step 1), long-term change during adulthood (step 2) plays an important role in community-level sound change (step 3).

1.2. Long-term. On a timescale of years, the traditional view is that an individual's linguistic system is largely fixed in adulthood (e.g. Chambers 2003:197). The assumption of accent stability over years underlies the widespread use of the apparent-time construct to study sound change in progress using a synchronic sample (Cukor-Avila & Bailey 2013).

Previous work has examined the extent of postadolescent accent plasticity in two kinds of settings where accent change toward changing community norms seems intuitively likely: individuals who remain in the same speech community, where some change is in progress (panel studies; Sankoff 2005, 2013), and individuals who have moved between dialect regions (dialect change studies; Auer et al. 2005, Siegel 2010).[1] For example, Sankoff and Blondeau (2007) examine change in rhotic realization ([r] → [ʀ], a community-level sound change in progress) in Montreal French using a panel of thirty-two speakers, recorded eleven years apart, and find relative stability in most individuals (72%), with the remainder increasing their [ʀ] use. While most large-scale dialect change studies do not break down results by individual speakers, those that do suggest that significant accent change in adulthood is uncommon. For example, Foreman (2003) examines six American speakers who settled in Australia for whom longitudinal data is available over ten to twenty-seven years; five of the six speakers show little or no significant change across six phonological variables. However, case studies have highlighted the fact that dramatic accent change is possible. For example, Queen Elizabeth II's realization of English vowels has changed significantly over fifty years of radio addresses, often paralleling community-level changes (Harrington et al. 2000, et seq.); and Yiddish folk singer Sarah Gorby shows a mixture of stability and change across phonological variables over fifty years, with change generally toward the standard (away from her native dialect) (Prince 1987). More generally, long-term studies find that both variables and speakers (for a given variable) differ substantially in the degree of plasticity over years (Siegel 2010:51), for a variety of reasons. Though further

[1] We do not consider age grading, where a variable's use changes during adulthood in a predictable way.

work is needed given this heterogeneity, the picture that emerges to date from long-term studies across a range of languages and variables (e.g. Bowie 2005, Brink & Lund 1975, Nahkola & Saanilahti 2004, Sancier & Fowler 1997, Stanford 2008) is that 'the default for adults is apparently stability', while a minority show significant pronunciation change, usually in the direction of community-level changes in progress or ambient norms (Sankoff 2013:274).

Generational change. The finding that adults show limited and heterogeneous long-term plasticity in pronunciation goes naturally with the view that change during adulthood is not a primary driver of community-level sound change. An influential proponent of this view is Labov (1994, 2007), who argues that pronunciation changes internal to speech communities occur primarily via generational change: intergenerational transmission of norms from adults to children, and incrementation by children in the direction of changes in progress during childhood and adolescence. By contrast, the typical mechanism of changes rooted in dialect contact is diffusion in the course of interaction between adults who use different pronunciations. Diffusion is less important than transmission as a source of linguistic diversity because it is driven by adults, who show less and more sporadic 'capacity to change their linguistic systems' compared to children and adolescents (Labov 2007:349). While Labov's account has been subject to debate (e.g. Babel et al. 2013), it represents a common view: change in adulthood is not an important factor in the major source of sound change (internal change), while external change is a less important source precisely because change in adulthood is limited and sporadic.

1.3. Medium term. Empirical evidence from the short-term and long-term literatures broadly suggests that accents are both plastic during interactions and largely stable over the lifespan, and associated theoretical viewpoints reach conflicting conclusions about accent plasticity in individuals and its relationship to community-level sound change. This conflict comes from extrapolating from short-term and long-term results to accent plasticity in individuals over the medium term (days–months), about which little is known.

Long-term studies suggest that accent stability is the norm in adulthood over years, but little is known about accent dynamics in adults on shorter timescales (cf. Barden & Großkopf 1998, Evans & Iverson 2007, Pardo et al. 2012, where sampling points are separated by months). A reasonable null hypothesis for medium-term dynamics would be that an individual's speech does not change over days–months; we call this the stationarity hypothesis. We can distinguish two versions of the stationarity hypothesis, thinking about the trajectory of an individual's 'baseline' use of a phonetic variable every day, over a period of months (schematized in Figure 1), after controlling for other factors (style, linguistic context, etc.). In the strong version, there could be no time dependence at all (type A), either from day to day or over longer periods. A weaker version would be that individuals vary from day to day (by-day variability), but not over longer periods (type C). There is almost no work examining whether an individual's accent fluctuates from day to day (cf. Heald 2012, Pisoni 1980, discussed below), though Nahkola and Saanilahti (2004) suggest this possibility.[2]

Short-term studies show that speakers regularly adjust their pronunciation due to the speech of others. The proposed link between short-term shifts in interaction and longer-term change is at the heart of the CBA model and motivates many short-term studies (e.g. Babel 2011, Delvaux & Soquet 2007, Nielsen 2011, Pardo 2006). An important initial assumption being made in these literatures, which we call the persistence hy-

[2] In contrast, Cukor-Avila (2015) and Rickford and Price (2013) have examined daily fluctuations in individuals' use of morphosyntactic variables, focusing on effects of style shifting and methodology.

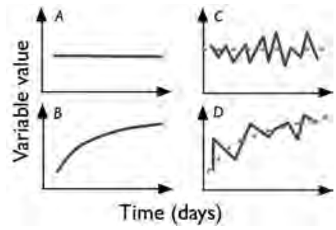

FIGURE 1. Schematic of possible types of time dependence in a phonetic variable within individual speakers: no change (A), time trend (B), by-day variability (C), by-day variability and time trend (D). Solid line indicates the variable's value on each day; dashed lines indicate its mean value over several days.

POTHESIS, is that short-term shifts persist beyond individual interactions; they may then accumulate and eventually lead to change in an individual's pronunciation norms over the medium term. To our knowledge, however, there is no work examining the persistence of short-term shifts in use of a phonetic or phonological variable for more than one hour.[3] The persistence hypothesis predicts that if an individual engages in the same type of interactions sufficiently often, there should be steady change in his 'baseline' use of a phonetic variable over days–months—beyond any by-day variability—which we term a TIME TREND. A phonetic variable could show a time trend without by-day variability (Fig. 1: type B), or both a time trend and by-day variability (Fig. 1: type D). By-day variability and time trends form two independent dimensions of medium-term time dependence: an individuals' pronunciation could show either, both, or neither.

1.4. MEDIUM-TERM ACCENT DYNAMICS IN A CLOSED SYSTEM. The disconnect between short-term and long-term plasticity and the paucity of empirical data on timescales in between motivates the current study, which investigates MEDIUM-TERM time dependence in speech production in individuals by taking advantage of a 'natural experiment': the reality television show *Big Brother UK*, whose structure is uniquely suited to investigating how and why an individual's accent changes over the medium term. The show contains speech from the same individuals, recorded on a near-daily basis over three months, making it possible to examine whether aspects of contestants' speech show by-day variability, time trends, or both. Contestants interact with each other constantly, without access to the outside world; the house is thus a linguistically 'closed system' (Bane et al. 2010), a small community where persistence of short-term shifts in pronunciation might be expected to be especially likely, and where it is in principle possible to test whether change in an individual's speech can be related to social interaction.

We first describe a corpus of spontaneous English speech from one season of this show, in which we examine five phonetic variables for twelve contestants over up to three months (§§2–3). For each variable, we build statistical models to characterize its time dependence within each speaker, after controlling for other factors (§4). We then use the results of these models to address two research questions about accent dynamics over the medium term.

First, what qualitative kinds of time dependence do phonetic variables show within individual speakers over three months (§5)? We consider the types of time dependence shown in Fig. 1, and we find that medium-term time dependence of one kind or another is ubiquitous across speakers and variables, and that time dependence is due primarily to by-day variability and secondarily to time trends. Second, to what extent can we ac-

[3] Goldinger (2000) shows persistence over one week of an overall percept, but acoustic variables are not examined.

count for the observed patterns of variability over time (§6)? We consider the same types of potential sources that mediate short-term pronunciation shifts—social factors, linguistic factors, and individual differences—and find some evidence for a role of each in shaping medium-term accent dynamics.

These findings allow us (§7) to evaluate the stationarity and persistence hypotheses, to situate medium-term dynamics between the disparate patterns individuals show on short-term and long-term timescales, and to evaluate the mechanism of accent dynamics in individuals and the possible relationship to community-level change.

2. The big brother corpus.

2.1. Show description. The Big Brother corpus consists of speech from the ninth season of the reality television show *Big Brother UK* (Channel 4/Endemol), which aired from June 5 to September 5, 2008 (ninety-three days).[4] Twenty-one contestants (or housemates) took part in the show: sixteen entered the Big Brother house on day 1, and five more entered later in the season. Housemates were gradually evicted during the season by a combination of nomination by other housemates and voting by the viewing public. The last housemate remaining won a cash prize.

While on the show, housemates had essentially no linguistic input from the outside world: they did not leave the house, interact with people not involved with the show, or (with rare exceptions) have access to media. They spent most of their time interacting with each other in some form, for example, in unstructured conversations or while participating in tasks set by Big Brother. The one important exception was the diary room, an isolated room where housemates could go to speak with 'Big Brother'. Big Brother was in fact different people (male and female) at different times; he/she could see housemates, but could not be seen by them, and communicated only through audio. Housemates went to the diary room to talk about their feelings, events in the house, and so forth, and could either go to the diary room voluntarily or be called by Big Brother.

Housemates were recorded at almost all times, including by wearable microphones. During the season, the public could see video of housemates via a live feed or various produced shows, including daily 'Highlights' shows and weekly 'Diary room uncut' shows, which consisted of continuous segments from the house, presented without commentary.

Table 1 gives basic demographic information for the twenty-one housemates, who come from diverse dialect regions. Sixteen housemates are native speakers of British dialects. Five housemates reside in the UK but are not native speakers of British dialects. Sara and Darnell are native English speakers from Australia and the USA, respectively; Sylvia and Mohamed were born in Sierra Leone and Somalia, respectively, and speak near-native English with light accents. Kathreya is a native Thai speaker whose English is heavily accented and frequently ungrammatical.

2.2. Corpus description. The corpus consists of all segments from the 'Highlights' and 'Diary room uncut' shows where a single housemate was in the diary room, which we call diary room clips (or clips). Audio from all clips was broadcast quality. The 749 clips contain roughly 14.5 hours of housemate speech. The corpus is unbalanced across housemates (Table 1), because housemates were on the show for different amounts of time: the less time a housemate spent in the house, the less frequently their speech was sampled (clips occur once per 0.8–5.4 days, for different speakers). Thus, to address our research questions about time dependence in individual housemates' use of

[4] Information about the show and housemates comes from Wikipedia (2012a,b). This corpus is an expanded version (roughly twice as large) of that used in Sonderegger 2012.

SPEAKER	GENDER	AGE	DIALECT REGION	DAYS ON SHOW	# OF CLIPS	HOUSEMATE SPEECH
Dale	M	21	N. England	65	43	0:43
Darnell	M	26	USA	93	73	1:24
Kathreya	F	31	L2 Thai	90	56	1:08
Lisa	F	40	N. England	86	53	1:02
Luke	M	21	N. England	58	49	1:09
Michael	M	34	S. Scotland	93	81	1:42
Mohamed	M	25	London	90	51	0:53
Rachel	F	24	S. Wales	93	69	1:03
Rebecca	F	21	W. Midlands	51	34	0:32
Rex	M	24	London	93	71	1:25
Sara	F	27	Australia	64	39	0:44
Stuart	M	25	N. England	57	32	0:38
Alexandra	F	23	London	14	17	0:25
Belinda	F	44	S. England	15	7	0:08
Dennis	M	24	S. Scotland	23	7	0:09
Jennifer	F	22	N. England	30	13	0:20
Mario	M	43	N. England	37	15	0:19
Maysoon	F	29	London	27	5	0:03
Nicole	F	19	S. England	22	19	0:24
Stephanie	F	19	N. England	9	9	0:08
Sylvia	F	22	London	23	6	0:09
TOTAL					*749*	*14:28*

TABLE 1. Demographic information, length of stay in the Big Brother house, and amount of data in the Big Brother corpus for each of the twenty-one housemates from *Big Brother 9 UK*. The horizontal line separates the twelve 'core housemates' (top) and the other housemates.

each phonetic variable—in particular, distinguishing between by-day variability and time trends—we can only consider housemates who spend a relatively long period in the house. Fifty days (of ninety-three) was chosen as an arbitrary cutoff, leaving twelve housemates (of twenty-one), who we refer to as the CORE HOUSEMATES (see Table 1). The core housemates account for 85.6% of speech in the corpus and are the focus of most analyses below.

The corpus is limited to diary room clips of single housemates in order to best address our research questions, given that it was not feasible to transcribe more than a fraction of the produced episodes (>100 hours). Limiting the corpus to one type of interaction in a relatively constant setting allows us to minimize differences in speaking style between different days (i.e. clips) and thus better assess the stationarity hypothesis.[5] The only interaction in the clips is with Big Brother, whose role is usually limited to brief questions or answers. The register of the speech is generally casual and conversational, with characteristics of a sociolinguistic interview, but with the clear self-awareness and performativity expected given that the interaction may eventually be televised. The corpus allows us to examine how each housemate's 'baseline' linguistic usage varies over time, abstracting away from short-term shifts that may occur during conversation with other housemates.

Each clip was segmented into speaker turns and orthographically transcribed by research assistants and the first author. The orthographic transcription and audio of each clip were then force-aligned using a version of the HTK-based aligner from FAVE (Rosenfelder et al. 2011, Young et al. 2006), customized for the Big Brother Corpus.

[5] Note that it was not possible to assess time of day (e.g. morning vs. evening) given information in the corpus; see §7.1 for discussion.

3. DATA. We examine the dynamics of five aspects of pronunciation for speakers in the Big Brother house (PHONETIC VARIABLES): voice onset time (VOT), coronal stop deletion (CSD) rate, and the quality of three vowels (GOOSE, STRUT, and TRAP, using the 'lexical set' notation of Wells 1982 for a vowel's realization in a given dialect).

These variables provide complementary evidence for our research questions about the existence and sources of medium-term time dependence in pronunciation. First, all variables differ greatly across varieties of English, meaning the possibility exists for contestants' pronunciations to influence each other. VOT has been found to be very flexible over short-term timescales (e.g. Nielsen 2011, Sancier & Fowler 1997), making it a logical place to look for medium-term plasticity. Examining CSD lets us study whether medium-term time change obtains for more categorical aspects of pronunciation, as well as continuous phonetic parameters. The three vowels examined differ along two dimensions that are possible sources of medium-term time dependence and might affect the likelihood of convergence between housemates. Previous work has argued that, compared to GOOSE, the realizations of STRUT and TRAP have high SOCIAL SALIENCE across the UK, which has been argued to constrain which variables can shift over the short and long term (e.g. Babel 2010, Trudgill 1986) and might affect a variable's medium-term plasticity as well. GOOSE is undergoing SOUND CHANGE in communities across the UK (e.g. Docherty 2010, Haddican et al. 2013, Hawkins & Midgley 2005), while STRUT and TRAP are relatively stable. Speakers might show more flexibility in the realization of a vowel undergoing sound change, if their exposure to a greater number of variants (outside the house) gives them a broader range of 'self-exemplars' to draw on in converging toward other speakers.

We first introduce each phonetic variable, then describe annotation and data-cleaning steps, and summarize the resulting data set.

3.1. VOICE ONSET TIME. VOT, the time difference between the onset of a stop's release burst and the onset of voicing in a following segment, is an important phonetic cue for the contrast between English 'voiced' and 'voiceless' stops (e.g. Docherty 1992, Lisker & Abramson 1967). VOT may be positive (burst + aspiration duration) or negative (duration of voicing preceding burst). (Phonologically) voiceless stops are produced with variable degrees of positive VOT, while (phonologically) voiced stops are produced with either shorter positive VOT or negative VOT. Variability in VOT has been studied mostly in lab speech (see Auzou et al. 2000, Docherty 1992), but some recent work examines spontaneous English speech (Chodroff et al. 2015, Yao 2009), including Stuart-Smith and colleagues' (2015) study of variability in positive VOT in spontaneous Glasgow vernacular—whose analytical choices we often follow, and who discuss the issues involved in measuring VOT in spontaneous speech. Like Stuart-Smith and colleagues, we measured positive VOT (summed burst and aspiration duration) for every WORD-INITIAL stop token produced with a burst or aspiration, without taking account of negative VOT or voicing during the closure. Thus, 'voiced stops' throughout this article always means PHONEMICALLY voiced stops, as opposed to stops produced with negative VOT.

ANNOTATION AND DATA SET. We measured VOT semi-automatically for all word-initial stops in the corpus (n = 16,784 voiced, 13,777 voiceless), using a procedure similar to that of Stuart-Smith et al. 2015.

First, an AUTOMATIC MEASUREMENT of VOT for each token was obtained by applying AutoVOT (Keshet et al. 2014, Sonderegger & Keshet 2012). AutoVOT requires a classifier that has been trained on manually labeled VOT data, as well as a window of time in the audio file for each token (specified in a Praat TextGrid; Boersma & Weenink 2011) in which to search for the beginning of the VOT interval. Applying the classifier to each token in a file yields predicted VOT intervals, which are outputted on a new tier of the TextGrid. To predict VOT for voiceless and voiced stops, we used the classifiers for voiceless and voiced English stops distributed with AutoVOT. The window for each token was taken to begin and end 25 ms before and after the force-aligned segment boundaries. The parameters for minimum/maximum predicted VOT were set to 15/250 ms for the voiceless classifier and to 5/150 ms for the voiced classifier, and other algorithm parameters were kept at default values.

Second, in the MANUAL CORRECTION phase, two or three phonetically trained annotators (two for voiced, three for voiceless stops) reviewed the predicted VOT intervals in Praat. For each interval, the annotator: (a) checked whether no burst was present (e.g. stop is realized as a fricative), in which case the token was marked for exclusion; (b) determined whether the VOT interval boundaries were where she would have placed them if annotating VOT manually; and (c) manually corrected the boundaries, if this was not the case. Step (a) resulted in 3,474 excluded tokens. Manual annotation (steps b–c) was performed in Praat. The left boundary was placed wherever the first 'large' amplitude increase in high-frequency frication occurred, established using the amplitude track and spectrogram; if there was a gradual rise in frication, the boundary was placed at the midpoint. When multiple clearly separate bursts occurred, the last one was used as the left boundary. The right boundary was determined primarily using the waveform: if periodicity was not present before the burst, at the zero crossing closest to the onset of periodicity; if periodicity was present throughout the burst, at the point where amplitude began to rise and general waveform shape changed abruptly.

A number of further exclusions were made. All 1,511 tokens from Kathreya were excluded, since her data may show extensive transfer from Thai (which has a three-way VOT contrast for stops). A total of 280 tokens with missing values for variables used in the models below (listed in Table 2) were excluded. Eighty tokens with VOTs outside of 1–80 ms for voiced stops, or 8–175 ms for voiceless stops, were excluded as having extreme VOTs, with the cutoffs determined by visual inspection of the distribution of VOT (separately for voiced and voiceless tokens). Because speech rate has a large effect on VOT relative to other variables, we excluded sixty-two tokens with a speech rate greater than ten syllables/second as having extreme speech rates (presumably due to forced-alignment errors), with the cutoff determined by visual inspection of the distribution of all tokens.

The final VOT data set consists of 12,908 voiced tokens (from 788 words, 678 clips) and 12,246 voiceless tokens (from 964 words, 668 clips), across twenty speakers.

3.2. CORONAL STOP DELETION. Coronal stop deletion (a.k.a. *t/d*-deletion) is a variable process in English in which word-final coronal stops (/t/ and /d/) are sometimes deleted in word-final consonant clusters (e.g. *best* as [bɛs] vs. [bɛst]).[6] CSD has been examined in dozens of studies (usually) of spontaneous speech across many varieties of English over the past fifty years (reviewed in Hazen 2011, Schreier 2005, Tagliamonte & Temple 2005). Fewer studies have examined CSD in British English varieties (e.g. Tagliamonte & Temple 2005, Temple 2009); particularly notable is Tagliamonte and Temple's study of York English, whose analytical choices we often follow. Most previous work treats

[6] CSD is one type of final consonant cluster reduction (CCR) in English and the most commonly studied. See Hazen 2011 on the relationship between CSD and CCR.

CSD as a binary variable, although 'deletion' is clearly a categorical approximation for gradient realization of coronals (e.g. Temple 2009). We use a categorical binary notion of 'deletion' in order to make use of the results of previous work—including Tanner et al. 2017, where the CSD data set described below is analyzed from a different perspective, unrelated to our research questions about time dependence.

ANNOTATION AND DATA SET. Annotation was carried out by four phonetically trained annotators, for all speakers except Kathreya (twenty speakers), whose data was excluded due to a near-categorical deletion rate (presumably because word-final clusters ending in *t/d* are phonotactically illegal in Thai). Each token (n = 14,259) for these twenty speakers whose underlying form ended in a *t/d*-final consonant cluster was manually annotated for the surface realization of the preceding consonant and the phonological context surrounding the final coronal, the length of any following pause (or other nonspeech), and the CELEX wordform ID (which was used to determine morphological class), as well as the realization of the final coronal.[7]

Coronal stop realization was first annotated at a fine-grained level, using spectral and auditory criteria. Annotators chose from eight categories describing possible surface realizations of *t/d* (using the acoustic cues of burst, glottalization, sudden closure; realization as a glottal stop, an unreleased coronal, etc.). In cases where the underlying *t/d* was followed by a surface *t/d* (e.g. *want to*), the target *t/d* was taken to be realized only in cases where there was evidence for realization of a sequence of two distinct coronal stops. If the second *t/d* was clearly present following a closure (usually due to the presence of a burst) and no other evidence of a previous *t/d* realization was present (such as a glottal stop), the *t/d* was taken to be the realization of the following word, and the word-final *t/d* was annotated as unrealized.

The fine-grained annotation was then collapsed into a binary variable of present versus deleted, following previous work (e.g. Tagliamonte & Temple 2005), where the presence of any unambiguous phonetic reflex of the underlying *t/d* indicates presence. We also followed Tagliamonte and Temple (2005) in discarding all tokens ending in /rt/ or /rd/ in rhotic varieties, given that many housemates are speakers of nonrhotic varieties.

We further excluded 141 tokens without a preceding consonant in their surface realization, twenty-two tokens where speech rate could not be reliably determined, one token with annotation errors, and fifty-nine tokens at turn ends. This last step was necessary in order to include pause duration and following context as INDEPENDENT factors affecting CSD rate, following Tanner et al. 2017, where the motivation for this choice (rather than coding pause as an alternative following context) is discussed.

The final data set consists of 12,788 tokens (from 570 words, 667 clips), across twenty speakers.

3.3. VOWEL FORMANTS. Accents of English differ primarily in their vowels (Wells 1982). Within a given variety of English, vowel quality is determined primarily by the first two formants (F1, F2). We consider only variation in F1 and F2, for GOOSE, STRUT, and TRAP.

The realization of these vowels varies systematically across English dialect regions; this variation will be important for assessing whether factors such as social salience affect medium-term time dependence. We describe the expected pronunciations of each

[7] We did not make exclusions such as frequent (e.g. *and*) or contracted (e.g. *don't*) forms, as in much previous CSD work, instead accounting for lexical differences using by-word random-effect terms.

vowel in the English dialect regions for each of the twelve core speakers (see Table 1 above)—Southern England (including London), Northern England, West Midlands, Southern Wales, Southern Scotland, General American, and General Australian (Beal 2004, Clark 2004, Cox & Palethorpe 2007, Ferragne & Pellegrino 2010, Penhallurick 2004, Stuart-Smith 2004, Watson 2007, Wells 1982)[8]—as well as the range of realizations that might be expected for Kathreya, based on studies of Thai learner English, or transfer from her L1 (e.g. Tsukada 2008). Figure 4 in §5.2 below shows the approximate correspondence between IPA symbols used here and location in F1/F2 space.

GOOSE is subject to different degrees of fronting ([u] ~ [ʉː] ~ [y]) across varieties of English. In Southern Scotland and General Australian, a fronted variant has long been the norm, while in Standard Southern British English (SSBE), Northern English, West Midlands English, and many American varieties, fronting is a change in progress (e.g. Docherty 2010, Haddican et al. 2013, Hawkins & Midgley 2005). An exception is Welsh English, where a back [u] is the norm. We thus expect GOOSE to be fronted to different degrees for all native speakers, with the exception of Rachel (from Wales). We also expect Kathreya to show a less-fronted GOOSE than native speakers (close to [u]).

STRUT realization is one of the most characteristic features dividing varieties from the North and South of England. STRUT is merged with FOOT (characteristically [ʊ]) in Northern English and is strongly socially marked (Wells 1982); in accents of the West Midlands, the two are variably merged, with variable realizations ([ʊ] ~ [ɤ] ~ [ə]). They are not merged in other English varieties represented in our data set, in which STRUT is realized as a lax mid-low vowel ([ʌ] ~ [ɐ] ~ [ə]). This is also the expectation for Kathreya.

TRAP is typically a front [æ] in General American and Australian. In the UK, accents other than SSBE considered here tend to use a more centralized pronunciation ([a]). Traditionally [æ] is used in SSBE, but younger speakers have shifted toward [a] (Hawkins & Midgley 2005). Kathreya is expected to realize TRAP somewhere along the [æ] ~ [a] continuum.

In terms of social salience, previous work has argued that STRUT is highly salient (especially within England), while TRAP has medium social salience, and GOOSE has low social salience, in the sense of different degrees of fronting (e.g. Haddican et al. 2013, Wells 1982).

ANNOTATION AND DATA SETS. The vowel data sets consist of data only from the twelve core speakers. We considered all tokens produced by these speakers of the target vowels, defined as those whose reference pronunciation was [u], [ʌ], or [æ] in a word's 'primary pronunciation' in CELEX (Baayen et al. 1996). To avoid heavily reduced vowels, we excluded tokens whose force-aligned durations were less than 30 ms, tokens from a list of highly frequent words (*and*, *um*, *just*, *uh*, *to*), and tokens for vowels whose reference pronunciation is unstressed in the FAVE pronunciation dictionary. This procedure resulted in 4,580 tokens for GOOSE, 6,323 for STRUT, and 7,903 for TRAP, for which we measured F1 and F2 semi-automatically.

In the AUTOMATIC MEASUREMENT step, we obtained F1 and F2 for each token, using the same customized version of the FAVE suite used for forced alignment (see §2.2). FAVE was set to use the 'faav' method to determine measurement points,[9] and to per-

[8] 'General American' is an approximation of the dialect of the single American speaker (Darnell), an African American who was born in the UK, grew up in St. Louis, Missouri, and moved back to the UK as an adult (Wikipedia 2012a).

[9] Usually, this corresponds to one-third of the vowel's duration—including for the three vowels considered here, except GOOSE after coronal consonants, for which the vowel beginning is used.

form 'remeasurement' within each file; other parameters were kept at default values. Following automatic measurement, F1 and F2 were Lobanov-normalized within each speaker (Lobanov 1971), using the speaker's entire automatically measured vowel space, to control for physiological differences between speakers. We use normalized formants in all analyses below, usually referred to simply as F1 and F2.

MANUAL CORRECTION was carried out by two phonetically trained annotators using FVR (Schwartz 2015), a graphical program for filtering and remeasuring vowel formant data. The goals of manual correction were to exclude tokens realized using a 'different' vowel from the target, including highly reduced tokens, and to verify and correct as necessary the formant measurements of the remaining tokens. All candidate tokens of a vowel from a given speaker were plotted in an F1/F2 display in which it was possible to hear each token's corresponding audio. Each token was examined, perceptually judged, and either excluded or accepted. Tokens were excluded that were (perceptually) deleted or heavily reduced; produced during yelling, disfluencies, or with heavy glottalization; or realized (perceptually) as anything not on a whitelist that included all expected realizations in dialects represented in the house. All remaining tokens were accepted. If an accepted token's formants were judged to be possibly mismeasured (judging by position in F1/F2 space, auditory judgment), they were checked in Praat and manually remeasured if necessary. For GOOSE, 1,716 tokens were excluded, 1,562 for STRUT, and 2,482 for TRAP.

Finally, the first author examined F1/F2 plots of each speaker's data for each vowel, looking for any remaining extreme outliers in formant measurements or words labeled with the wrong lexical class. This led to the exclusion of thirteen GOOSE, nine STRUT, and thirty-six TRAP tokens.

The final vowel formant data set consists of F1 and F2 measurements from twelve speakers for 2,847 tokens for GOOSE (from 200 words, 588 clips), 4,743 for STRUT (from 345 words, 591 clips), and 5,365 for TRAP (from 434 words, 615 clips), across twelve speakers.

3.4. STATIC FACTORS. In building models of time dependence for each phonetic variable below, we control for a variety of factors besides time that affect phonetic realization (linguistic factors, social factors, properties of the utterance), which we call STATIC FACTORS, such as speech rate and the identity of surrounding segments. Table 2 summarizes the static factors for each phonetic variable. Because their effects are not directly related to our research questions about time dependence, previous work on static factors and how we control for them statistically are described in Appendix A.

4. ANALYSIS. From an analysis perspective, it is easiest to divide the data sets of phonetic variables just described into nine NARROW VARIABLES, each corresponding to a single phonetic parameter that may behave independently of the others: VOT for voiced and voiceless stops (two), CSD (one), and (normalized) F1 and F2 for each vowel (six). We now describe the analysis carried out on the data for each narrow variable, culminating in one DYNAMIC MODEL per speaker per narrow variable, describing time dependence in the variable during the speaker's time in the Big Brother house.

4.1. METHOD. The analysis of each narrow variable is conceptually similar. For each speaker, we fit four statistical models of the speaker's realization of the variable, each assuming one of the types of time dependence shown in Fig. 1, while controlling for static factors in the same way. These models are then compared in order to evaluate which qualitative type of time dependence best characterizes this variable, for this

VARIABLE	STATIC FACTOR	TYPE
VOT	Place of articulation	factor (bilabial, coronal, velar)
	Following segment type	factor (vowel, sonorant)
	Following vowel	factor (nonhigh, high)
	Syllable stress	factor (unstressed, stressed)
	Phrase position	factor (phrase-initial, phrase-medial)
	Speech rate deviation	continuous
	Speech rate mean	continuous
	Word frequency	continuous
	Gender	factor (male, female)
CSD RATE	Following context	factor (vowel, other consonant, *t*/*d*)
	Preceding context	factor (other obstruents, sonorants, sibilants)
	Pause duration	continuous
	Morphological class	factor (regular past, irregular past, nonpast)
	Word frequency	continuous
	Speech rate	continuous
	Gender	factor (male, female)
VOWEL FORMANTS	Preceding segment	factor (oral apical, nasal apical, oral labial, nasal, labial, liquid, obstruent + liquid cluster, palatal, velar, *w*/*y*, other)
	Following segment manner	factor (affricate, fricative, lateral, nasal, rhotic, stop, other)
	Following segment place	factor (alveolar, bilabial, interdental, labiodental, palatal, velar, other)
	Following segment voicing	factor (voiced, unvoiced, other)
	Vowel duration	continuous

TABLE 2. Summary of static factors included in the static models of each phonetic variable.

speaker. The best model is chosen, resulting in one model of time dependence per variable, per speaker. For example, we fit four models of VOT for voiceless stops for Rachel, controlling for the static factors listed in Table 2: one model assumes no systematic time dependence in VOT; one assumes some time trend (mean VOT changes over time) but no additional by-day variability; one assumes by-day variability in VOT around an unchanging mean value (no time trend); and one assumes both by-day variability and a time trend.

At a technical level, the analysis of each narrow variable uses both mixed-effects regression models (MEMs; e.g. Baayen et al. 2008, Gelman & Hill 2007) and generalized additive mixed models (GAMMs; Hastie & Tibshirani 1990, Wood 2006), fitted using the lme4 and mgcv packages in R (Bates et al. 2014, R Core Team 2014, Wood 2011). An MEM is first fitted for each variable for data from ALL speakers, as a function of static factors only. These STATIC MODELS allow us to determine, with maximum statistical power, which static factors should be controlled for in modeling time dependence within INDIVIDUAL speakers. The results of the static models serve as input to GAMMs modeling the time dependence of each variable within individual speakers, which we refer to as DYNAMIC MODELS.

Before describing the modeling procedure for each variable, we discuss our choice of this particular analysis method and then briefly introduce GAMMs.

CHOICE OF ANALYSIS METHOD. Some discussion of our analysis method is warranted, as splitting up the data by speaker and dividing the analysis into two steps both carry risks. For our data, we would ideally build a single model of both static factors and time for each phonetic variable, across all speakers. This was not possible because speakers turn out to show qualitatively different patterns of time dependence, making it necessary to consider four possible types of time dependence (Fig. 1) for each speaker. Finding the best possible combination of types of time dependence for all twelve speakers within a single model would require comparing prohibitively many models (4^{12}). Thus, we decided to compare possible models of time dependence within each speaker separately, for a given variable.

We considered two ways of doing this: (i) building four GAMM models accounting for both static factors and time dependence within an individual speaker, and (ii) building four GAMM models of time dependence for an individual speaker using the residuals from a static model fit to all speakers. Method 1 has the disadvantage of less accurate and lower-power estimates of the effects of static factors, for which the data from ALL speakers is relevant; this could lead to controlling less accurately for the effects of static factors, and possibly underestimating time dependence within a speaker as a result. We use method 1 for the CSD data, where method 2 is conceptually impossible (as discussed below). Method 2 has disadvantages associated with any case of 'residual regression', where the effect of one set of predictors X_1 (here, static factors) on a dependent variable Y is 'regressed out', and then the residuals are modeled as a function of another set of predictors X_2 (here, time)—as opposed to building a single model, $Y \sim X_1 + X_2$. Residual regression can lead to bias in the estimated effect of X_2, which can be either conservative or anti-conservative depending on factors such as the correlation between X_1 and X_2 (Darlington & Smulders 2001, Freckleton 2002). Despite these issues, residual regression is widely used in language research—for example, to control for the effect of word length on reading times in self-paced reading studies (e.g. Ferreira & Clifton 1986, Fine et al. 2013), to remove by-speaker and by-item variation in acoustic parameters before examining the effects of predictors of interest on prosody (Breen et al. 2010), or in short-term and long-term studies analyzing change over time in VOT within speakers after controlling for other factors (Stuart-Smith et al. 2015, Yu et al. 2013). Here, we use method 2 for the VOT and vowel formant data, in order to control for static factors with maximum statistical power.

GENERALIZED ADDITIVE MIXED MODELS. While MEMs are now widely used in language research, GAMMs may be less familiar (for introductions see Winter & Wieling 2016, Wood 2006). GAMMs are an extension of MEMs: the dependent variable can still depend on factors and continuous variables (fixed-effect terms), which may vary across groups (random-effect terms), but it can now also depend on a SMOOTH of one or more independent variables: a nonparametric function that locally fits the data, conceptually similar to a nonlinear smoother. GAMMs are closely related to GENERALIZED ADDITIVE MODELS (GAMs), which similarly extend nonmixed regression models (e.g. linear and logistic regression) by allowing for smooth terms. Applications of these methods to linguistic data, in order to model variability across time and space, include event-related potentials (Kryuchkova et al. 2012), dialect variation (Wieling et al. 2014), and language evolution experiments (Winter & Wieling 2016). The crucial aspect of GAMMs for our purposes is their ability to model largely arbitrary patterns of variability over time—here, in a phonetic parameter.

GAMMs allow for incorporation of two types of terms, which conceptually correspond to the two types of time dependence that differentiate the four patterns in Fig. 1: (i) a random intercept of CLIP, which captures by-day variability (because different clips for a given speaker generally occur on different days); (ii) a smooth of DAY, which captures any time trend. Thus, we build four models per speaker/variable pair: one without (i) and (ii), one with term (i) only, one with term (ii) only, and one with both terms (i) and (ii). These models are compared using the Akaike information criterion (AIC), a widely used model-selection criterion.[10] The model with the lowest AIC is chosen, resulting in a single model of time dependence for a given narrow variable per speaker, after controlling for static factors.[11]

4.2. Building dynamic models. We now summarize how the analysis was carried out for each variable, using the data sets described in §3.

Voice onset time. The dependence of VOT on static factors was first determined by fitting two linear mixed-effects models of log(VOT) to the VOT data from the twenty speakers, for voiced and voiceless stops. These static models included (i) fixed-effect terms for all static factors (Table 2), including interactions based on exploratory data analysis, as well as annotator identity; and (ii) near-maximal by-speaker and by-word random effects (Barr et al. 2013). The model-fitting procedure and results are detailed in Appendix B1.

We then built dynamic models of time dependence of VOT within the eleven core speakers, controlling for static factors, by modeling the residuals of the static models of VOT, offset for each speaker by the static model's estimate of their 'mean VOT' (the model intercept, plus the speaker's fitted random intercept). These offset residuals, which we call residualized VOT, have the interpretation for each token of 'VOT, after controlling for static factors'.[12] For each speaker, for each type of stop (voiced, voiceless), we fit four dynamic models of residualized VOT, corresponding to the four possible types of time dependence. Each model is a GAMM with a Gaussian link and includes a by-word random intercept, as well as appropriate smooth terms (as described above). The model with the lowest AIC was selected from each set of four models. This procedure resulted in twenty-two models of time dependence (11 speakers × {voiced, voiceless}) of VOT within individual speakers.

Coronal stop deletion. Tanner and colleagues (2017) fit a mixed-effects logistic regression model of CSD realization for the data set described above from twenty speakers. We use the results of this static model only to motivate the static factor terms included in the dynamic models of CSD (see below), rather than its residuals serving as input to the dynamic models (as for VOT); thus, only aspects of the model relevant for the dynamic models are summarized here. As fixed effects, the model included main-

[10] Best practices for model selection for GAM(M)s are not settled, especially with respect to selecting random-effect and smooth terms, as in the models compared here, for which various statistical issues arise (see Pinheiro & Bates 2000, Wood 2006). Comparison via AIC is one of several possible model-selection procedures. A referee recommends model comparison via a chi-squared test on the difference in ML score, which can be more robust than the AIC but cannot be applied when comparing nonnested models (here, corresponding to types B and C of time dependence). We use AIC based on the discussion in the gam.selection page of Wood 2015, but note that methodology for model selection for longitudinal phonetic data is an important issue for future work.

[11] An important consideration in regression modeling of longitudinal data is the possibility of autocorrelated residuals due to nonindependence of consecutive time points (see e.g. Pinheiro & Bates 2000:Ch. 5); this often occurs for linguistic data from laboratory experiments (van Rij et al. 2016). Examining residual autocorrelation is not straightforward for our data due to its complex temporal structure (irregularly spaced time points, many tokens per time point, i.e. per clip). Nonetheless, as a basic check, we examined autocorrelation plots of simplified versions of the predictions of each dynamic model and found no evidence for significant autocorrelation of residuals, possibly because measurement points are spaced much further apart (days) than in experimental data.

[12] More precisely: after controlling for static factors, as well as variability among words, since the residuals factor out by-word random effects.

effect terms for all static factors (Table 2), as well as a term for annotator identity, and other main-effect and interaction terms motivated by the research questions of Tanner et al. 2017. Each variable was coded such that the model intercept has the interpretation of a grand mean (and this coding was retained for the dynamic models): factors were coded using Helmert contrasts, and continuous variables were standardized. Prior to standardizing, 0.01 seconds were added to pause duration, and both pause duration and speech rate were log-transformed. Main-effect terms for six static factors—following and preceding context, word frequency, speech rate, pause duration, and annotator identity—reached significance, as well as several interaction terms, of which pause duration-by-following context had a much larger effect size than other interactions.

We built dynamic models of time dependence of CSD rate for the eleven speakers (twelve core speakers, minus Kathreya), after controlling for these seven terms. Unlike for VOT, the residuals of the static model of CSD realization do not have the interpretation of 'variable value, after controlling for static factors', which cannot be calculated for individual tokens for a binary variable (see e.g. Agresti 2002:§6.2). In order to model time dependence in CSD rate for a given speaker while controlling for static factors, we instead built models for that speaker's data including terms for both time dependence and static factors. For each speaker, we fit four dynamic models of CSD realization, each one a GAMM with a logit link, including smooth terms capturing time dependence. Each model included the seven terms for static factors and a by-word random intercept.[13] For each speaker, the model with the lowest AIC was selected, resulting in eleven models of time dependence (one per speaker).

VOWEL FORMANTS. Six linear mixed-effects models of normalized formants were first fitted to the data from the twelve speakers for which vowel formant data was obtained—two models (for normalized F1, F2) for each of GOOSE, STRUT, and TRAP. These static models included terms for all static factors (Table 2), parametrized in a different way from the VOT and CSD models, due to the relative sparsity of data (fewer tokens; see Appendix B2). The models included (i) fixed-effect terms for the effects of following consonant voicing and vowel duration; (ii) three random intercept terms for the effects of following consonant place, following consonant manner, and preceding segment class; (iii) by-speaker and by-word random intercepts; and (iv) a by-speaker random slope for vowel duration. The model-fitting procedure and results are detailed in Appendix B2.

We then built dynamic models of time dependence in F1 and F2 within the twelve core speakers, after accounting for static factors, in a manner similar to VOT, by modeling the residuals of the static models, offset for each speaker by the static model's estimate of their mean formant value. These RESIDUALIZED FORMANTS have the interpretation within each speaker of 'F1 for GOOSE, after controlling for static factors' (etc.). For each speaker, for each variable, the model with the lowest AIC was selected, resulting in seventy-two models of time dependence (12 speakers × {F1, F2} × {GOOSE, STRUT, TRAP}) of vowel formants within individual speakers.

[13] Static factors were coded in the GAMMs as in the regression model of Tanner et al. 2017, with two exceptions. The effect of speech rate was modeled using a smooth term (as for Day), based on exploratory plots suggesting that speakers showed qualitatively different speech-rate effects. Pause duration was modeled as a log-transformed continuous variable, rather than discretized into four levels as in Tanner et al. 2017, which was motivated by the research questions of that study.

4.3. Summary. The procedure just described resulted in 105 dynamic models (22/11/72 for VOT/CSD/vowels) describing the dynamics of each phonetic variable in each core speaker (with the exception of VOT and CSD for Kathreya, the nonnative speaker). We present and discuss the results of these models in the next two sections, corresponding to our two research questions: what time dependence does pronunciation show within individual speakers over days–months (§5), and to what extent can we explain these patterns (§6)?

5. Time dependence. In this section, we describe each models' predictions of the existence and nature of time dependence in each variable. For each variable, we plot the models' predicted patterns of time dependence, in order to visualize the extent of time dependence beyond any effect of static factors or random noise.[14] For each variable, we also summarize time dependence in terms of the four-way categorization discussed above (Fig. 1). A more detailed summarization of the predicted patterns of time dependence for each variable for each speaker, including quantitative measures, is given in the online supplements (§S1),[15] but is not assumed here.

5.1. Voice onset time. Figure 2 shows the model-predicted time dependence in residualized VOT for voiced and voiceless stops for each speaker, which we call predicted VOT. Because of how residualized VOT was computed, and the coding of static factors included in the static models (see Appendix B1), predicted VOT can be interpreted as a speaker's 'baseline VOT' on a given day (averaging across words, spoken at the speaker's average speech rate, etc.). The dashed line in each panel shows the predicted time trend: the mean predicted VOT on a given day, without taking by-day variability into account. A flat dashed line indicates no time trend. When a model predicts by-day variability, the estimated magnitude of these daily fluctuations is expressed by a fitted parameter in the model, denoted σ, such that roughly 95% of days (i.e. clips) are predicted to have log(VOT) within $\pm 2\sigma$ of the mean.[16] The amount of by-day variability is shown in each panel by shading: the vertical range of shading shows the predicted range of VOT between a 'high day' and a 'low day' ($\pm 2\sigma$ from the mean, represented by the dashed line). Finally, it is also possible to extract from the model an estimate of the size of a fluctuation in VOT on a given day (the best linear unbiased predictor; Pinheiro & Bates 2000:§2.2). Combining this estimate with the predicted mean gives the solid line, which shows the actual predicted VOT on a given day. The qualitative type of time dependence in VOT for each speaker is summarized in Table 3 below (columns 2–3).

Results. All speakers show some time dependence in VOT, for both voiced and voiceless stops: by-day variability, a time trend, or both. In particular, every speaker shows by-day variability in VOT, for both voiced and voiceless stops. Of the eleven speakers, six (54.5%) show a time trend in predicted VOT. This corresponds to changes in both voiced and voiceless stops for four speakers, and only in voiceless stops for two speakers.

5.2. Coronal stop deletion. Figure 3 shows the model-predicted time dependence in CSD rate for each speaker (analogously to Fig. 2), holding all static factors at their

[14] We plot model predictions rather than empirical means in all figures showing time dependence (Figs. 2, 3, 5–8) in this article because these predictions essentially are the results of the models that are relevant for our research questions—which the figures report in graphical form. (Unlike regression models such as MEMs, GAMMs do not include interpretable numerical coefficients for smooths or random-effect terms.) Plots based on empirical means generally look very similar, but are not reported here due to space constraints.

[15] The online supplemental materials are available at http://muse.jhu.edu/resolve/24.

[16] Formally, σ is the variance component for the by-clip random intercept.

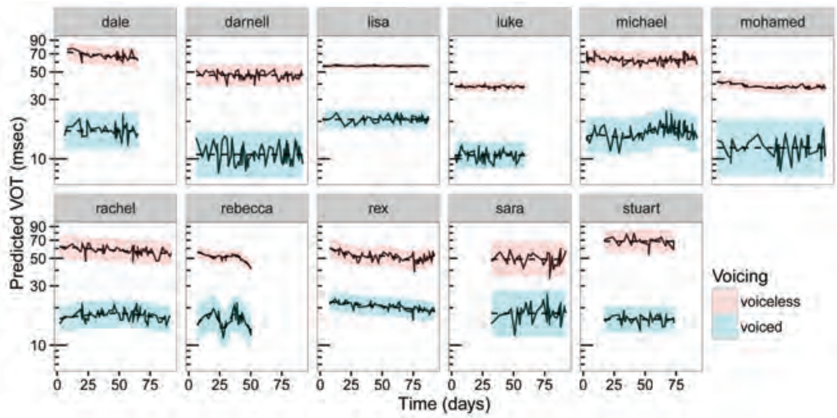

FIGURE 2. Predicted time dependence of VOT (in ms, on log scale) for each speaker, for voiced and voiceless stops. Solid lines indicate model predictions for each day, including by-day variability. Dashed lines indicate predicted time trend of mean VOT (without any by-day variability). Shading indicates magnitude of by-day variability, between 'low' and 'high' days. (See text.)

average values and ignoring by-word random effects; the predictions can be interpreted as 'average CSD rate' (for an average word, at a speaker's mean rate, etc.). In each panel, the dashed line shows the predicted time trend in mean CSD rate (if any). The magnitude of fluctuations in CSD rate is parametrized by σ such that roughly 95% of days are predicted to have a CSD rate (expressed in log-odds) within $\pm 2\sigma$ of the mean. Shading shows the amount of by-day variability in CSD rate ($\pm 2\sigma$ around the mean), and the solid line shows the actual model-predicted CSD rate on each day. The qualitative type of time dependence in CSD rate for each speaker is summarized in Table 3 (column 4).

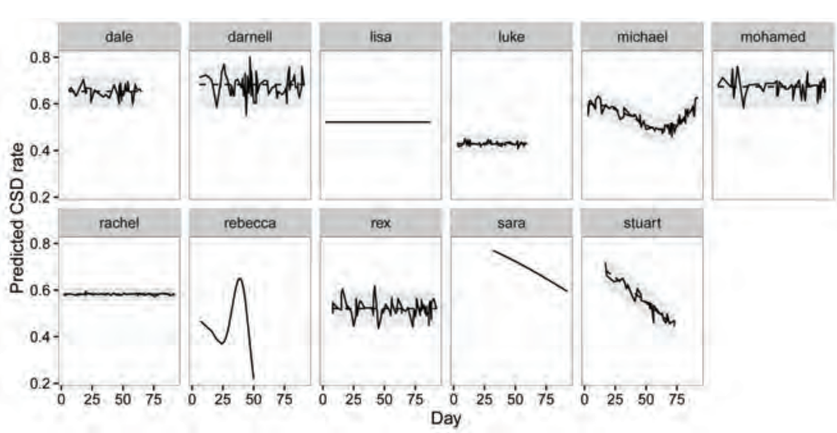

FIGURE 3. Predicted time dependence of CSD rate for each speaker. Solid/dashed lines and shading are defined as in Fig. 2.

RESULTS. Ten of the eleven speakers (91%) show some time dependence in CSD rate: by-day variability (six), a time trend (two), or both (two). Eight speakers (73%) show by-day variability. Four of eleven speakers show time trends.

5.3. VOWEL FORMANTS. Before turning to the results of the dynamic models for the vocalic variables, we examine the model-predicted mean locations of each vowel: how

SPEAKER	VOT		CSD RATE	GOOSE		STRUT		TRAP	
	VOICED	VOICELESS		F1	F2	F1	F2	F1	F2
Dale	C	D	C	D	C	C	D	C	D
Darnell	C	C	C	C	D	D	C	C	D
Kathreya	—	—	—	C	C	D	D	D	D
Lisa	C	C	A	C	A	C	C	C	D
Luke	C	C	C	C	C	A	D	C	D
Michael	D	D	D	D	C	C	D	C	D
Mohamed	C	D	C	C	C	C	D	C	C
Rachel	D	D	C	C	D	C	C	D	C
Rebecca	D	D	B	C	D	C	D	D	C
Rex	D	D	C	C	C	C	C	C	D
Sara	C	C	B	D	C	C	C	D	C
Stuart	C	C	D	B	C	D	C	C	A

TABLE 3. Summary of the qualitative type of time dependence each speaker shows for each variable, using notation from Fig. 1: no change (A), time trend (B), by-day variability (C), by-day variability and time trend (D).

each speaker realizes GOOSE, STRUT, and TRAP, on average, after controlling for static factors (i.e. residualized F1 and F2). How speakers differ in realizing these vowels will be important in interpreting time dependence in F1 and F2 below. Figure 4 shows each speaker's predicted mean F1 and F2 for each vowel, with superimposed IPA symbols showing the approximate phonetic realization corresponding to different parts of the F1/F2 space.

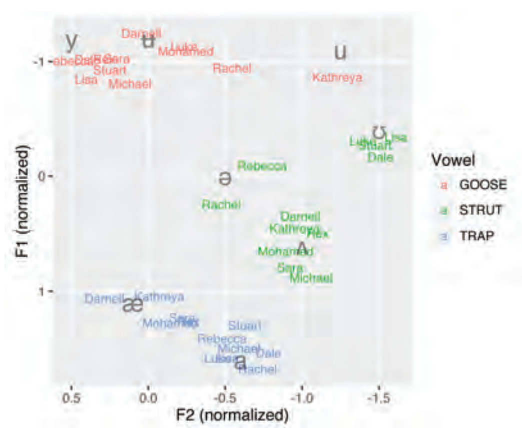

FIGURE 4. Predicted mean values of F1 and F2 (normalized) for GOOSE, STRUT, and TRAP, for each speaker (shown using names). Superimposed IPA symbols show approximate locations of the main phonetic variants used for these vowels across dialects (see §3.3).

For GOOSE, the main division is between Kathreya, who shows a very backed [u], and the native speakers (all others). Among the native speakers, Rachel shows the most backed realization, as expected for a Welsh English speaker. All of the other speakers have markedly fronted GOOSE, to varying degrees. These realizations are all consistent with our expectations (§3.3).

For STRUT, we can distinguish three groups of speakers. Dale, Lisa, Luke, and Stuart—the speakers from Northern England—have markedly higher and backer STRUT than the other speakers, as expected given that they have merged STRUT and FOOT (usually realized [ʊ]). Rachel and Rebecca realize STRUT as more centralized on average than the other speakers, while the last group—Darnell, Kathreya, Michael, Mohamed, Rex, Sara—realize STRUT as more low and back, something like [ʌ].

For TRAP, the housemates vary roughly along a line from [æ] to [a]. We can nonetheless distinguish three groups for the purposes of exposition: Darnell and Kathreya, Sara/Rex/Mohamed, and all of the others (Dale, Lisa, Luke, Michael, Rachel, Rebecca, Stuart). The third group consists of speakers from regions where the norm is [a], all of whom indeed realize TRAP closer to [a] than [æ]. In the second group, Mohamed's and Rex's intermediate realizations are expected given their dialect area (Southern England), while Sara's realization is unexpectedly far from [æ]. In the first group, Darnell's TRAP is clearly [æ], as expected for an American speaker; Kathreya's realization is similar, within the range of realizations expected for Thai-accented English.

In sum, the housemates' mean realizations of the three vowels fit our expectations, given each speaker's dialect background (§3.3).

TIME DEPENDENCE. We now turn to the dynamic models. Figure 5 shows the model-predicted time dependence in F1 and F2 for each speaker for each vowel, which we call PREDICTED F1/F2. Analogously to the VOT and CSD cases, predicted F1 and F2 can be interpreted as a speaker's 'baseline F1 and F2' on a given day (at an average vowel duration, and averaging over possible types of preceding and following contexts). In each panel, the dashed line shows the predicted time trend in mean F1 and F2, vertical shading illustrates the magnitude of by-day variability between 'high days' and 'low days' (again characterized by the parameter σ, such that 95% of days have formant values within $\pm 2\sigma$ of the mean), and solid lines show model-predicted formant values on each day. The qualitative type of time dependence in F1 and F2 for each vowel for each speaker is summarized in Table 3 above (columns 5–10).

RESULTS. All of the speakers show some time dependence in the realization of all three vowels (taking F1 and F2 together). In particular, all show by-day variability in the realization of all three vowels: either F1, F2, or both are predicted to fluctuate from day to day around a mean value. For the vast majority of speaker/vowel pairs (thirty-two of thirty-six), both F1 and F2 show by-day variability. A subset of speakers show time trends in the realization of each vowel (in at least one of F1 and F2) during the season: seven speakers for GOOSE (58%), eight speakers for STRUT (67%), and ten speakers for TRAP (83%).

5.4. SUMMARY. We can make several observations about medium-term time dependence in pronunciation based on Table 4, which summarizes this section's results.

First, across all variables, time dependence is ubiquitous: in all but one case (sixty-eight cases; 98.5%), a speaker's pronunciation fluctuated daily, showed steady change in mean realization over time, or both. Medium-term time dependence takes on very heterogeneous forms across speakers and variables, but the EXISTENCE of some time dependence is constant. Second, by-day variability in pronunciation is nearly ubiquitous (95.65% of cases). In particular, for every gradient variable (VOT, vowel formants), every speaker showed daily fluctuations in pronunciation; the two cases without by-day variability are for CSD. Third, time trends are often present, but are markedly less common than by-day variability (43.48% of cases), both within and across variables. For each variable, some speakers did not show steady change in mean pronunciation over time—at least not change that had an effect size large enough for us to detect.

6. SOURCES OF TIME DEPENDENCE. So far, we have found that time dependence in phonetic variables over days–months is near-ubiquitous. The observed patterns of time dependence are also very heterogeneous (Figs. 2, 3, 5). We now turn to our second

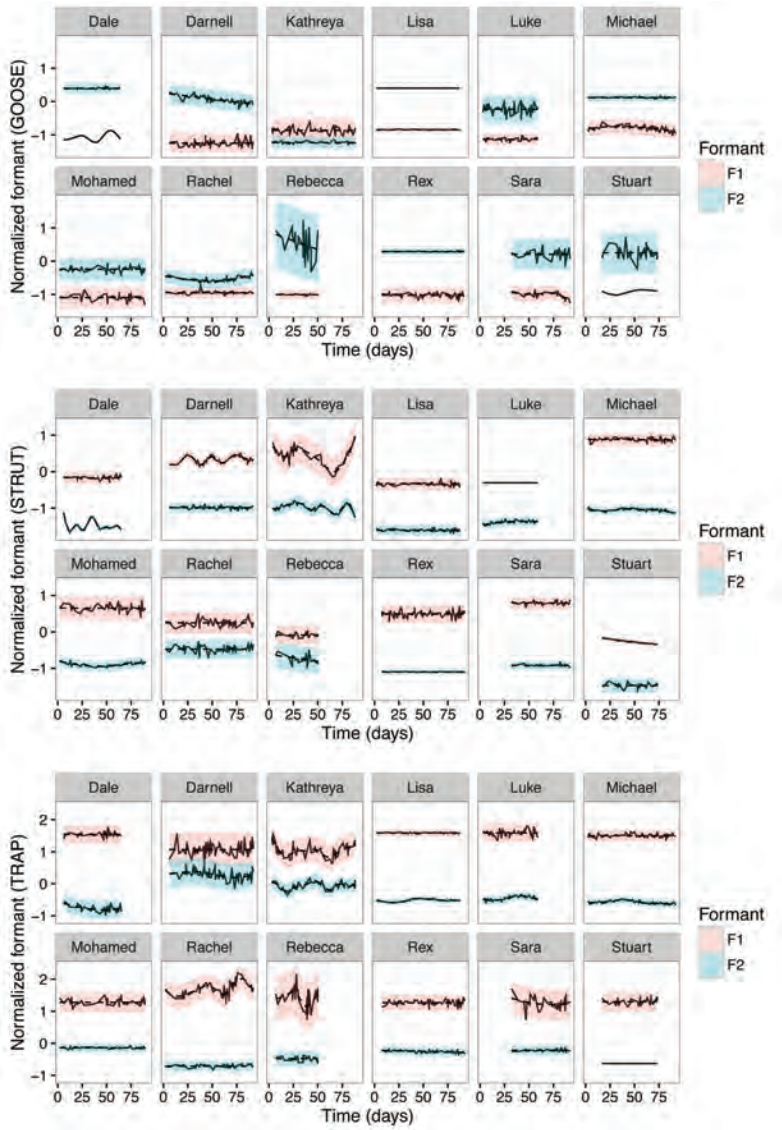

FIGURE 5. Predicted time dependence of (normalized) F1 and F2, for each speaker, for GOOSE, STRUT, and TRAP. Solid/dashed lines and shading are defined as in Fig. 2.

VARIABLE	TYPE OF TIME DEPENDENCE			
	NONE (A)	TT ONLY (B)	BDV ONLY (C)	BDV & TT (D)
VOT (voiced)	0	0	7	4
VOT (voiceless)	0	0	5	6
CSD	1	2	6	2
GOOSE	0	0	5	7
STRUT	0	0	4	8
TRAP	0	0	2	10

TABLE 4. Number of speakers showing each type of time dependence (Fig. 1), for each variable: none, time trend only, by-day variability only, or by-day variability and time trend.

question: can we account for any of the variability in patterns of medium-term time dependence? Short-term time dependence in phonetic variables is influenced by linguistic factors, social factors, and individual differences (§1.1), suggesting that medium-term time dependence may be as well. We consider three possible sources of time dependence of these types: CONTRAST between sounds corresponding to different variables; CONVERGENCE in pronunciation of individual variables; and differences in SPEAKER PLASTICITY, across all variables.

6.1. CONTRAST. The contrastive role played by phonetic variables, including vowel formants (e.g. Babel 2011), has been invoked to explain the results of short-term experiments where they are examined (Mitterer & Ernestus 2008, Nielsen 2011). Here, we consider one way in which variability over time in vowel formants might be constrained by their use as acoustic cues differentiating speech sounds: variability in a speaker's realization of a sound might be constrained by the variability in its realization across English dialects.

The realization of English vowels differs greatly among English varieties, including those represented in the house. If a speaker's realization of a vowel changes enough over time, it could be realized as a variant not usually associated with the speaker's variety. To examine whether this ever occurs, Figure 6 shows ellipses describing the range of each speaker's predicted production during the season, for each vowel (ellipses contain 95% of predicted F1/F2 values, across clips), which we discuss with respect to the groups of speakers identified above for each vowel, according to their realization in the corpus and speakers' dialect regions (§§3.3, 5.3).

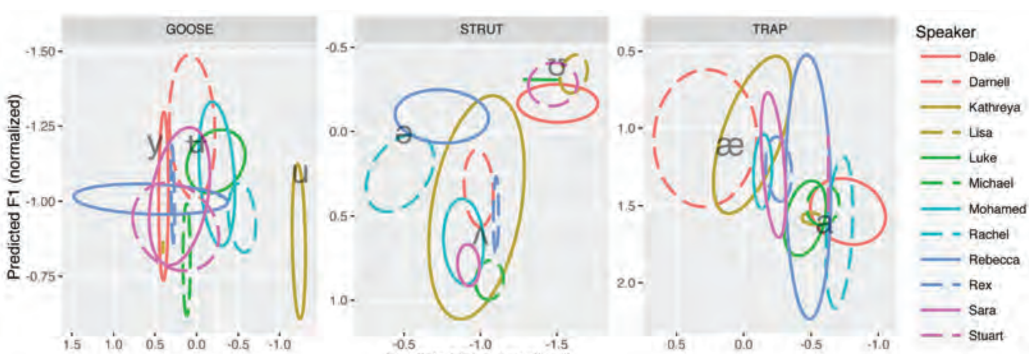

FIGURE 6. 95% confidence ellipses for predicted realizations of GOOSE, STRUT, and TRAP over time for each speaker. Superimposed IPA symbols show approximate locations of the main phonetic variants used for these vowels across dialects.

For GOOSE, the primary division was between Kathreya (clearly backed [u]) and the other speakers (different degrees of fronting). This division is maintained over time. The secondary division, among the native speakers only (excluding Kathreya), was between Rachel (who showed the least fronting) and the others. This division is largely maintained: although Luke's GOOSE is also notably backed, on most days Rachel remains the native speaker with the most backed variant.

For STRUT, the primary division was between speakers from Northern England and all others. This division is maintained over time: the ellipses for the Northerners (Dale, Lisa, Luke, Stuart) and for the other speakers do not overlap: the former always use something like [ʊ], while the latter never do. A second division, among non-Northerners, was between speakers who use variants closer to [ə] (Rachel, Rebecca) or closer to [ʌ]. Setting aside Kathreya, the [ə] and [ʌ] groups of native speakers retain distinct realizations of STRUT over time.

For TRAP, speakers showed a continuum of realizations from [a] to [æ], and we distinguished three groups for convenience. Speakers' ranges of realizations of TRAP over time show more overlap, compared to GOOSE and STRUT. However, the three GROUPS of speakers largely do not overlap: again setting aside Kathreya, there is little overlap between the ellipses for Darnell, Mohamed/Rex/Sara, and all of the other native speakers.

Thus, speakers do not generally use a markedly different phonetic variant for a given vowel over time, relative to the variants used in dialects represented in the house.

6.2. CONVERGENCE. Convergence could take the form of converging time trends for a phonetic variable among all speakers ('overall convergence') or a subset of speakers ('partial convergence'). We examine the evidence for each kind of convergence using Figures 7–8, which show the predicted time trajectories of each variable, WITHOUT taking by-day variability into account. These trajectories are the time trends for speakers whose dynamic models contain time trends, and otherwise a flat line at the speaker's mean value.

OVERALL CONVERGENCE. The fact that only some speakers show time trends for each variable makes the strongest form of overall convergence impossible: speakers cannot all shift their mean pronunciations toward each other over weeks–months, since some of them do not shift at all over this timescale. However, it is possible that we have missed some time trends due to a lack of statistical power. We can still ask whether observed time trends are at least consistent with overall convergence, by asking for each variable whether time trends that DO occur move toward each other.

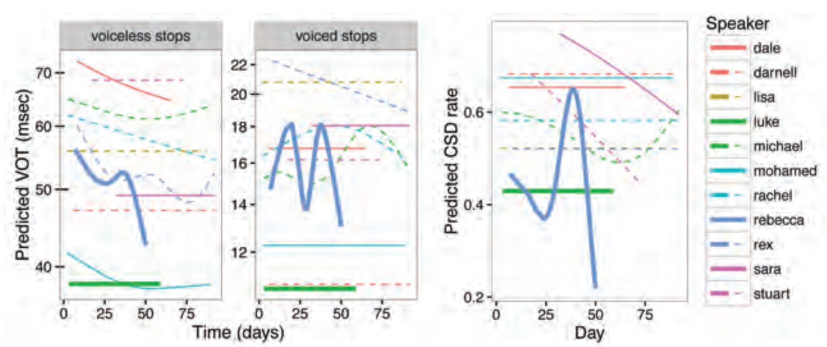

FIGURE 7. Trajectories of predicted mean VOT (left; ms, log scale) and mean CSD rate (right), over time, for each speaker. Bolded lines correspond to Luke and Rebecca, two speakers who may show convergence.

Neither VOT (voiced or voiceless stops) nor CSD rate shows clear evidence for overall convergence: in Fig. 7, speakers with time trends do not systematically converge toward each other over the course of the season. The same is true in most cases for the vowels (Fig. 8): F1 and F2 of STRUT and TRAP, and F1 of GOOSE. In the case of F2 for GOOSE, there is possible evidence for convergence: from about day 50 on, the three speakers who do show time trends move toward a similar value. This pattern must be considered tentative, however, as it is based on only three speakers.

PARTIAL CONVERGENCE. Partial convergence is more difficult to assess, because of the possibility of spurious results. For each variable, there are certainly pairs of housemates whose time trends seem to move toward each other over the course of the season (e.g. Michael and Sara for CSD). However, such pairs would also exist if speakers showed random time trends in phonetic parameters. The most solid evidence for partial convergence would be observing the same pattern of convergence across several pho-

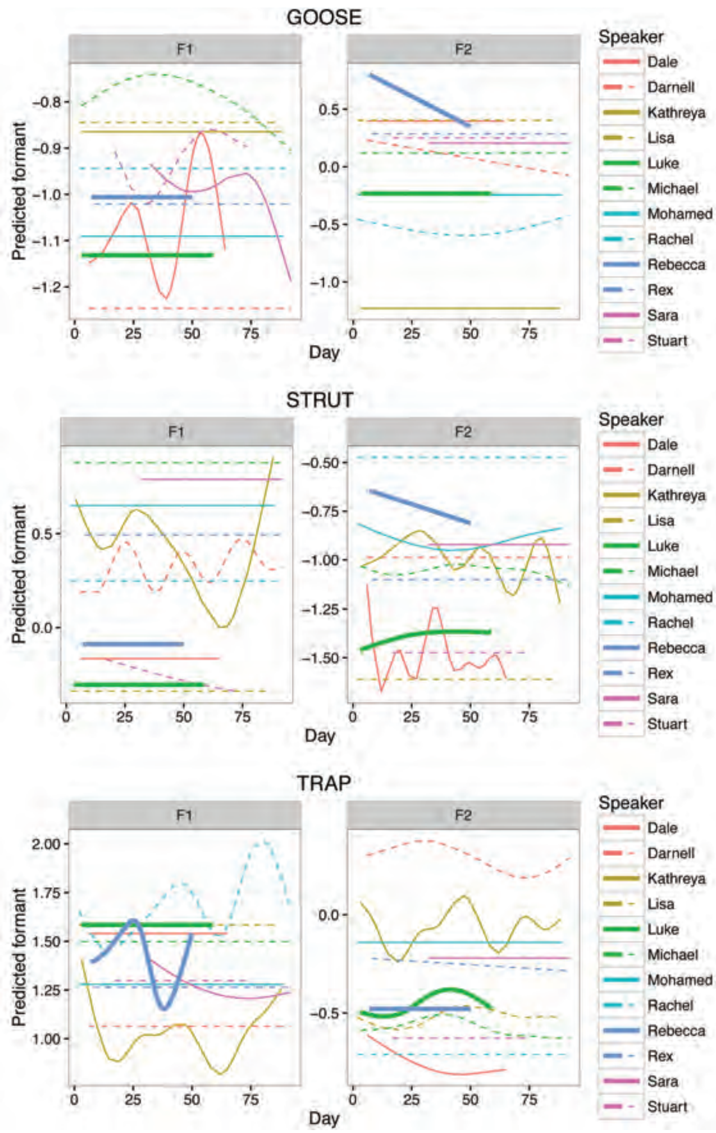

FIGURE 8. Trajectories of predicted mean (Lobanov-normalized) F1 and F2 over time, for each speaker, for GOOSE, STRUT, and TRAP. Bolded lines correspond to Luke and Rebecca, two speakers who may show convergence.

netic variables, for the same subset of housemates, for whom there is a plausible reason for convergence to occur. We consider one such case.

Which pair of housemates would be most likely to converge? Following different theories of how short-term shifts may accumulate into accent change (§1.1), we might expect the most convergence between housemates who interact most often with each other, or who feel most positively toward each other. Both expectations arguably point toward the same housemates: Luke and Rebecca. These two housemates quickly become friends after entering the house and interact progressively more with each other over the course of the season, until Rebecca is evicted on day 51. During this time, they develop romantic interest in each other, but at first only discuss this with other housemates. On day 37, they admit their mutual interest to each other, and afterward form an inseparable couple. Luke and Rebecca's bond appears to have been genuine: after the show ended, they moved in together and remained in a relationship for over a year. No other pair of housemates on the show formed such a bond during the season or interacted with each other as often, making Luke and Rebecca seem the most likely pair of housemates to converge.

To assess whether they do, the bolded lines in Figs. 7–8 above plot Luke's and Rebecca's predicted mean values for each variable over time. Rebecca's GOOSE is progressively less fronted, moving steadily toward Luke's norm. For STRUT, Luke's [ʊ]-like production and Rebecca's centralized production move slightly, but steadily, toward each other over the season. For the remaining variables, the pair's pronunciations generally become more similar after day 37 or so: Rebecca's VOT shifts downward toward Luke's value, her pronunciation of TRAP shifts (back) toward his norm, and her CSD rate shifts radically downward, reaching and then surpassing Luke's low value. On the whole, the pair's realizations of phonetic variables become markedly more similar over time—especially during their period of most intense interaction, after their relationship qualitatively shifts (days 37–51)—suggesting convergence.

6.3. SPEAKER PLASTICITY. Anecdotally, our results so far suggest another possible factor influencing patterns of time dependence: speakers may differ in how variable their accents are, across variables (PHONETIC PLASTICITY)—whether this variability is due to a specific cause (e.g. convergence) or to random variation.

The clearest example is Lisa, who shows strikingly little variability over time compared to the other speakers. For vowels, this is visually clear from Fig. 6; she is also the only speaker to show no time dependence in CSD rate (Fig. 3), and she shows little time dependence in VOT (Fig. 2). Luke and Rebecca provide another example: it is primarily time trends in Rebecca's usage that are responsible for the pair's convergence across variables, and the magnitude of variability over time generally seems greater for Rebecca than Luke, across variables (Figs. 2, 3, 5). Both observations suggest Rebecca's speech may be systematically more variable over time than Luke's.

CALCULATING PLASTICITY OVER TIME. To test whether there are systematic individual differences in phonetic plasticity, we compute a single measure of plasticity per variable, per speaker, which takes both by-day variability and time trends into account, for each of six variables: VOT (voiced stops), VOT (voiceless stops), CSD rate, and the realization of GOOSE, STRUT, and TRAP. We calculate these measures using the 'residualized' predictions of the dynamic models (residualized VOT, etc.), which describe time dependence in each speaker's use of a phonetic variable after controlling for static factors. We compute a measure of dispersion for each variable for each speaker, as follows:

- VOT: For each VOT subset (voiced, voiceless), for each speaker, we take the standard deviation of the predicted log(VOT) value for each of the speaker's clips.
- CSD RATE: For each speaker, we take the standard deviation of the predicted log-odds of deletion for each of the speaker's clips.
- VOWELS: For each vowel, for each speaker, we first extract the predicted F1 and F2 for each clip. We then compute the centroid of these values (the center of the ellipses shown in Fig. 6) and the Euclidean distance of each clip from the centroid (in F1/F2 space), and take the average of these distances.

This procedure results in six values per speaker (see Table S4 in the online supplemental materials) that describe the degree of variability in their realization of each variable over the season, which we call PLASTICITY values.

Testing for differences in plasticity. To test whether speakers differ significantly in plasticity, across all variables, we carry out a permutation test (see e.g. Good 2013). The test statistic is the mean of rank correlations (Kendall's τ) between pairs of plasticity values, which we denote as μ.[17] The observed value of μ, denoted μ_{obs}, is 0.265, suggesting a weak tendency for speakers to have similar plasticity across variables. The permutation test assesses how likely it is that we would observe a value of μ this large, under the null hypothesis that each speaker's degree of plasticity is randomly distributed across speakers, independently for each phonetic variable (i.e. speaker plasticities for different variables are unrelated).

We carry out this test by (a) randomly permuting plasticity values among speakers, independently for each phonetic variable; (b) recalculating μ; and (c) repeating steps (a) and (b) $n = 100{,}000$ times. The n values of μ approximate its distribution under the null hypothesis. The p-value for this test is the probability (using this distribution) that $\mu \geq \mu_{obs}$, approximated as the proportion of the n runs for which $\mu \geq \mu_{obs}$. In our case, $p = 2.8 \times 10^{-4}$, suggesting that the null hypothesis is very unlikely to be true, and thus that a speaker's plasticity values for different variables are positively related. That is, there are 'more plastic' and 'less plastic' speakers.

Lisa	1.33	Dale	6.17	Sara	8.17
Luke	4.17	Rachel	6.17	Darnell	8.33
Michael	4.50	Mohamed	6.83	Kathreya	9.33
Rex	4.83	Stuart	6.83	Rebecca	10.00

Table 5. Overall phonetic plasticity for each speaker. (See text.)

We quantify each speaker's degree of plasticity by their mean plasticity value (rank), across variables (Table 5). As expected, Lisa is the least plastic speaker, while Luke is much less plastic than Rebecca, who is the most plastic speaker.

6.4. Summary. Our findings in this section suggest that linguistic factors and individual differences account for some of the observed patterns of pronunciation change, and they suggest a possible role for social factors. For vowels, housemates generally do not produce a markedly different phonetic variant for GOOSE, STRUT, and TRAP over time, with respect to the possibilities represented in the house. There are marked individual differences in phonetic plasticity, across variables. There was little evidence for convergence across all housemates in any phonetic variable, with one possible exception. We did, however, observe convergence across variables between a pair of housemates with an unusually close bond. Social factors may also help explain patterns of variability in vocalic variables, as discussed below.

7. Discussion. Our results on medium-term dynamics in phonetic variables in the Big Brother house allow us to address the questions raised in the introduction: what trajectories do phonetic variables show within individual speakers over three months, and how do the results bear on the persistence and stationarity hypotheses? How do medium-term accent dynamics fit in between short-term dynamics and long-term dynamics, and how does the answer bear on the striking discrepancy between the plasticity of accents on these two timescales? What factors affect medium-term accent dynamics in individuals, and what is the relationship of these dynamics to the mechanism of community-level sound change?

[17] These correlations are shown in Table S5 in the online supplemental materials.

7.1. Medium-term plasticity. Medium-term time dependence was ubiquitous: in 98.5% of cases, a speaker's pronunciation of a variable varied systematically over time. Fluctuations between recording sessions on different days were the norm (by-day variability; 95.65% of cases); steady change over longer timescales was less common, but still occurred often (time trends; 43.48% of cases). Our results suggest that medium-term variability in pronunciation over time is due primarily to short-term fluctuations (days), and secondarily to longer-term change (weeks–months).

Stationarity and persistence. Our results allow us to assess the stability of phonetic variables within individual speakers over days–months, about which little is known. Two reasonable null hypotheses, based on extending the stability many speakers show over years to shorter timescales, were that speakers do not systematically vary over time or over a timescale longer than days (the 'strong' and 'weak' stationarity hypotheses).

Pronunciation is almost never static from day to day, suggesting that the strong stationarity hypothesis can be rejected. Previous sociolinguistic work (Nahkola & Saanilahti 2004) has suspected the existence of such variability between recording sessions, but we are aware of only two laboratory studies examining variability within the same speaker's speech over days (Heald 2012, Pisoni 1980); both find no evidence for by-day variability in productions of English vowels by a small number of speakers. Our contrasting result could be due to our larger sample size or to greater variability in spontaneous versus laboratory speech.

An important caveat to our finding of ubiquitous variability across recording sessions is that some of this variation may not be a function of time per se (whether due to accumulated short-term shifts or to cognitive factors such as fatigue; Heald 2012). We have treated a housemate's diary room clips (speech to Big Brother) as representing 'baseline' use of each variable in the same social context, abstracting away from short-term shifts that may occur during interaction with other housemates. However, there could still be variability in speaking style between clips, due, for example, to style shifting or accommodation to different speakers acting as 'Big Brother'. Such stylistic and interviewer/interlocutor effects are of concern in all linguistic studies using data from multiple time points (e.g. Gregersen, Jensen, & Pharao 2017) and indeed are sometimes the focus (e.g. Gregersen, Jørgensen, et al. 2017, Rickford & Price 2013). Like most panel and dialect change studies, we take the existence of a systematic pattern of time dependence—in our case, the near-ubiquitous existence of by-day variability—to reflect at least in part change over time, rather than factors not controlled for. More work is needed using speech from a controlled setting to establish how large daily fluctuations are for phonetic variables when factors such as style shifting are controlled for. Further work could also better establish the timescale of variability. While our data from diary room clips (which largely occur on different days, within a given speaker) lets us establish variability over time on two timescales (time trends, fluctuations between clips), it does not let us distinguish between by-day variability and shorter timescales. It is possible that what we are calling 'by-day variability' actually reflects fluctuation on a scale of hours or minutes, instead of or in addition to daily variability. This possibility does not affect our main conclusions below, but accent dynamics on multiple timescales is an interesting direction for future work.

Pronunciation is also not generally static over weeks–months, suggesting that the weak stationarity hypothesis can be rejected. The existence of time trends is notable, because it offers less ambiguous evidence that speakers' 'baseline' use of a variable changes over time, compared to by-day variability. Time trends are unlikely to be due to style shifting or accommodation to different Big Brothers, since a speaker would need to incrementally shift in the same direction over many clips. Indeed, it is hard to think of reasons why a speaker's accent would systematically change over time—except for long-term accommodation to other housemates' speech. Thus, the existence of time trends suggests that MEDIUM-TERM PERSISTENCE of accommodation effects is possible, if very sporadic, across speakers and variables. More speculatively, it is also possible that some of the by-day variability in pronunciation is due to persistent accommodation effects.

7.2. THE RELATIONSHIP OF MEDIUM TERM TO SHORT AND LONG TERM. We now turn to the relationship between medium-term accent dynamics and shifts on shorter and longer timescales, in order to understand how individuals vary on different timescales.

In short-term studies, the evidence for the existence of shifts in pronunciation on a timescale of seconds–hours is relatively robust across speakers and variables, and the magnitude and direction of these shifts are heavily modulated by social factors, linguistic factors, and individual differences. In our data, time dependence on a timescale of days is near-universal, across speakers and variables. We argue that medium-term time dependence in our data is primarily due to BY-DAY VARIABILITY, and it can be partially accounted for by social factors, linguistic factors, and individual differences. In this sense, daily fluctuations in pronunciation (in our data) look qualitatively similar to shorter-term shifts (in previous work). By contrast, long-term studies suggest that accent stability is the norm in adulthood over a timescale of years, while a minority of adults show significant accent change. In our data, on a timescale of weeks–months, TIME TRENDS occur more sporadically than by-day variability: across all speakers and variables, there is steady change in pronunciation in less than half of cases. In this sense, accent change over weeks–months (in our data) looks qualitatively similar to accent change over years (in previous work). The similarity of different aspects of our results to the divergent results of the short-term and long-term literatures suggests that the disconnect between short-term and long-term accent plasticity is already present on a timescale of months: plasticity is widespread on timescales of up to days, but more sporadic over longer periods.

Why these similarities? We cannot say for sure, since we do not have data from the Big Brother house on short-term shifts (i.e. in interactions) or change over years. Nonetheless, we can sketch a possible account of the relationship between accent plasticity in individuals on different timescales, which addresses the mismatch between the ubiquity of short-term accent plasticity, on the one hand, and the heterogeneity of long-term accent change, on the other.

In the course of INTERACTIONS, some short-term shift nearly always occurs. We assume that at least some of the ubiquitous by-day variability in our data can be ascribed to accumulated short-term shifts—that is, that speakers 'bounce around' in their use of each phonetic variable as a result of short-term shifts in conversation, and thus have different 'baseline' values when recorded in a constant environment (the diary room) on different days. From DAY TO DAY, short-term shifts for a particular variable and speaker usually build up enough to 'change' her pronunciation. The amount of build-up is highly variable, due to all of the factors modulating short-term shifts. Next, note that in the vast majority of cases (for a given variable and speaker), the existence of a time trend implies the existence of by-day variability (Table 3). We hypothesize a relationship between these two types of time dependence: time trends are at least in part the result of accumulated by-day variability, but this build-up often does not occur (cases where there is no time trend) or is too small to be detected. The sporadic nature of change over weeks–months might be because speakers differ in how phonetically 'plastic' they are (see below); this could be capturing individual differences in the size of short-term shifts, or how systematically they accumulate. Regardless of the source, the result is that over WEEKS–MONTHS, accumulation of day-to-day shifts into longer-term change in pronunciation norms occurs for only a minority of speakers and variables—although day-to-day shifts COULD accumulate into change over weeks–months in pronunciation norms, this occurs only in a minority of cases. Accent change over YEARS is then also (correctly) predicted to be very heterogeneous.

CHANGE BY ACCOMMODATION. Our medium-term results bear on long-standing CBA theories, which propose a link between the short term and long term: short-term shifts in conversation (step 1) can accumulate into long-term accent change in individuals (step 2) and eventually community-level sound change (step 3). CBA theories have been difficult to test given the complexity of real-world speech communities and a lack of empirical data on medium-term dynamics (between steps 1 and 2); our medium-term results from a linguistically closed system provide one of the first (indirect) tests.

The simplest version of the CBA model is that accent change and sound change are primarily driven by 'automatic' convergence during conversation: the more interactions take place in a group, the more similar group members' speech should become. This idea, which has been repeatedly proposed (e.g. Bloomfield 1933, Delvaux & Soquet 2007, Paul 1880, Trudgill 2004), makes sense as a null hypothesis (Labov 2000:506). Our results suggest that this null hypothesis can be rejected: we see little evidence for overall convergence in how a set of housemates from diverse dialect regions speak over three months, despite constant interaction. This (non)finding bears most directly on steps 1–2 of the CBA model, which address accent dynamics in individuals. The combination of widespread medium-term time dependence in speakers' accents with a lack of overall accent convergence is consistent with the heterogeneous results of longitudinal studies examining accent change over years, for example, in college roommates or people who move between dialect regions (Barden & Großkopf 1998, Evans & Iverson 2007, Pardo et al. 2012, Wilson 2010). Both our results and previous work suggest that the dynamics of phonetic variables over time within individual speakers, even in settings of intense social contact, are highly complex. More empirical data on longitudinal variation within individuals is needed before the intuitively plausible link between social interaction and accent change during adulthood can be established, and it is also necessary for clarifying the nature of this link.

More complex versions of the CBA model hold that accent change and sound change are mediated by a variety of factors, because short-term shifts (step 1) are also mediated by these factors (e.g. Auer & Hinskens 2005, Garrett & Johnson 2013). Our results, and the account sketched above of the link between dynamics on different timescales in individuals, are consistent with this position: the heterogeneous patterns of medium-term time dependence in phonetic variables we observe can be partially explained in terms of social factors, linguistic factors, and individual differences—the same types of factors that modulate the direction and extent of shifts observed in short-term studies.

7.3. INDIVIDUAL ACCENT DYNAMICS AND COMMUNITY-LEVEL CHANGE. Medium-term time dependence in phonetic variables is both ubiquitous and heterogeneous, across speakers and variables. Our results allow us to assess several possible sources of accent dynamics in individuals—social factors, linguistic factors, and individual differences—that bear on the relationship between short-term shifts and longer-term change.

VOWEL DYNAMICS: CONTRAST AND SOCIAL SALIENCE. Our results for the vocalic variables suggest that time dependence in phonetic variables is constrained. For GOOSE, STRUT, and TRAP, despite extensive time dependence in their pronunciations, housemates generally do not produce markedly different phonetic variants with respect to the possibilities represented in the house, instead staying within the phonetic bounds expected given their dialect region. We suggest that the repertoire of possible realizations constrains variability in vowel realization: speakers generally do not produce variants that would 'sound different', for linguistic or social reasons, an interpretation similar to the role of SALIENCE invoked in both the short-term and long-term literatures (e.g. Auer et al. 1998, Kerswill & Williams 2002, Trudgill 1986).

Time dependence in vowel pronunciation may be constrained by contrast between speakers' preexisting phonetic categories. Findings from the short-term literature suggest that speakers' productions in imitation tasks (e.g. Babel 2011, Kim et al. 2011, MacLeod 2012) or auditory feedback tasks (Katseff & Houde 2008, Katseff et al. 2012) are strongly constrained by their existing 'phonetic repertoire' (or 'self-exemplars'; Garrett & Johnson 2013). The observed pattern of a speaker's vowel productions staying near a 'characteristic pronunciation' over days–months could simply result from using his phonetic repertoire over this timescale.

Social salience may also constrain variability over time, if speakers avoid productions that would 'sound different' in the sense of having social salience. GOOSE shows extensive overlap in how fronted each native speaker's realization is over time, while TRAP and STRUT show less overlap. GOOSE is fronting over time in many British varieties, and previous work has argued that this change has low social salience. Native speakers will thus have exposure to very different degrees of fronting and will not differentiate between degrees of fronting for social reasons; they are thus free to vary in GOOSE realization over time, as long as a fronted variant is used. This is exactly what is observed. STRUT realization is highly socially/regionally marked (with [ʊ]-like realizations strongly associated with Northern England). Thus, native speakers cannot vary their production into a region of F1/F2 space associated with a different variant without producing a socially salient difference; this constrained variation is what is observed. Finally, TRAP realization has medium social salience; this could constrain speakers' pronunciations to not vary too greatly along the [a] ~ [æ] continuum, as is observed. In sum, it is possible that social salience constrains what medium-term change in pronunciation is possible in a variable, echoing short-term and long-term studies that invoke social salience to explain which variables shift more or less (e.g. Auer et al. 1998, Babel 2010, Kerswill & Williams 2002, Kim et al. 2011, MacLeod 2012, Trudgill 1986). This possibility could be more rigorously tested in future work by using a larger number of variables and a sample of speakers who are more likely to share evaluative norms, such as economic migrants or university students who move from one dialect region to another (Barden & Großkopf 1998, Evans & Iverson 2007, Wilson 2010).

CONVERGENCE. In their simplest forms, both the CBA model and communication accommodation theory predict that the more a set of speakers interacts, the more similar their pronunciation norms will become (Auer & Hinskens 2005, Pardo et al. 2012). The Big Brother house seems like an ideal setting to observe accent convergence: constant interaction between speakers with very different accents, in a linguistically closed system. However, we found strikingly little evidence for OVERALL CONVERGENCE among housemates for any of the five phonetic variables examined, in the sense of converging time trends. In only one case (F2 of GOOSE) was there a pattern that could tentatively be interpreted as convergence, with the caveat that nine of twelve speakers showed no change in mean F2 over time. However, we did find clear PARTIAL CONVERGENCE for one pair of housemates, across all five phonetic variables. These results are consistent with a recent study of five pairs of college roommates by Pardo and colleagues (2012)—similarly motivated by the idea that intense interaction should lead to phonetic convergence—who found that 'overall levels of detected convergence … was [sic] modest, even after 3.5 months of relatively continuous cohabitation' (Pardo et al. 2012:197), with a single pair showing much clearer convergence than the others.

Why might we not have found overall convergence? The simplest possibility would be that convergence takes longer than three months (the duration of the show). Indeed, longitudinal studies of individuals who move between dialect regions have found that accent change is more widespread after one to two years than after three months (Barden & Großkopf 1998, Evans & Iverson 2007). We did find likely convergence for the pair of housemates who were arguably most likely to converge (Luke and Rebecca)—as measured by either frequency of interaction or social affinity—so convergence between housemates was possible over the timescale of the show. Why was it not more widespread? Luke and Rebecca may have simply interacted more frequently than the other housemates, but this cannot be the whole story: they interacted frequently for fifty-one days, but it is only in the last couple of weeks that convergence across variables is clear. We propose that convergence between them was also driven by their extremely strong social bond, cemented over those two weeks when they entered into a romantic relationship. Indeed, communication accommodation theory (Giles et al. 1991, et seq.) would predict Luke and Rebecca to express their increased emotional affinity toward each other in this period by converging during conversations, and these shifts to accumulate over time. Other housemates, who do not have as close a bond, do not accommodate toward each other as consistently, hence the lack of overall convergence. In short, accent change over three months may require social motivation, rather than just frequent interactions.

Another reason why we might not observe overall convergence is the performative aspect of the show. Because housemates are under constant observation, the show entails an extreme version of the 'observer paradox' (Labov 1972:113). Housemates are in a permanent, performative interaction with viewers, in which context they may (un)consciously manipulate their accent. Between different clips, phonetic variables may undergo short-term shifts as part of a speaker projecting a certain image (Le Page & Tabouret-Keller 1985) or depending on who the perceived audience is (Bell 1984). These shifts may eventually change an individual's pronunciation norms, as predicted by Auer and Hinskens's (2005) 'identity projection model', a CBA-like theory with the crucial difference that short-term shifts in conversation result from the speaker expressing a particular persona by shifting his usage toward what he thinks that persona sounds like, not relative to the interlocutor's speech per se. This source of time dependence may be unusually strong in the context of a televised show, swamping 'overall convergence' that might otherwise occur. (For example, housemates may explicitly want to sound distinct from other housemates, as part of their performed persona.) If this is the case, overall convergence would be more likely to be observed in other close-knit temporary communities where speakers are not constantly recorded, such as a summer camp.

Individual differences. We found strong evidence for one additional source of accent dynamics: housemates differ in overall 'phonetic plasticity', showing a characteristic degree of variability over time across phonetic variables. To our knowledge, the quantitative demonstration that some speakers are more variable than others across different aspects of pronunciation is novel, in line with a long-observed fact about interspeaker variation by sociolinguists—that some speakers are 'more variable' than others, in the sense of having a greater range of styles and variable realization (e.g. Eckert 2000, Rickford & Price 2013). Housemates' different degrees of plasticity could reflect individual differences in the cognitive mechanisms of perception and production, in the magnitude of style shifting, or something else. Establishing the extent and mechanism of individual differences in phonetic plasticity is an interesting direction for future work.

Regardless of their source(s), individual differences in phonetic plasticity may be relevant for community-level sound change. Any community-level change must be actuated by some speaker(s) deviating from the norm, then picked up by other speakers; the puzzle is why this happens at some times and not others, given that most sound changes are rooted in presumably universal phonetic-bias factors (Garrett & Johnson 2013). Much previous work suggests traits that make some people, who Milroy and Milroy (1985) call INNOVATORS and EARLY ADOPTERS, more likely to innovate and adopt new variants: social factors, including the range of variants individuals are exposed to due to life circumstances (Labov 2000); and lower-level individual differences, such as in cognitive processing style (Yu 2013) or the magnitude of coarticulation (Baker et al. 2011). Individual differences in phonetic plasticity may be another such factor. More plastic speakers are more likely to randomly generate new variants over time, making them more likely innovators (ceterus paribus). Thus, the more plastic an individual's speech, the more likely she is to use a variant that could seed a sound change in an interaction with an 'early adopter'. While speculative, this account adds to the list of systematic reasons why some individuals may be more likely sources of community-level sound changes.

8. CONCLUSION. We have used the 'natural experiment' of a reality television show to conduct the first systematic study of medium-term accent dynamics, over days to months, filling in an empirical gap between studies of short-term shifts during conversation and long-term accent plasticity over years. We considered five phonetic variables in a corpus of 14.5 hours of spontaneous speech, and built statistical models of time dependence in these variables within each of twelve speakers over three months, after controlling for a range of confounds. Our main findings are that time dependence in an individual's pronunciation over the medium term is near-ubiquitous and is primarily due to daily fluctuations, although systematic change over weeks to months occurs frequently as well. The combination of ubiquitous daily variability with more sporadic longer-term change mirrors the different extents of accent plasticity previously observed in conversation versus over the lifespan, suggesting that speakers show progressively less accent flexibility over progressively longer timescales. One question for future work is whether other aspects of an individual's linguistic system (e.g. lexical choice, morphology, syntax) show similar medium-term dynamics, as might be expected given that pronunciation is thought to be among the LEAST flexible aspects of an individual's linguistic system during adulthood (Sankoff 2005, Siegel 2010).

We argued that the heterogeneous patterns of medium-term time dependence in our data can in part be accounted for by the same kinds of factors that influence the degree and magnitude of short-term shifts in speech production: linguistic factors, social factors, and individual differences. Speakers systematically produce phonetic variants for vowels that remain within the boundaries expected given their dialect regions. We found evidence for accent convergence between two speakers with an exceptionally close social bond, but no strong evidence for accent convergence across all speakers. Finally, speakers show characteristic degrees of 'plasticity' over time across phonetic variables; more plastic speakers may be more likely to seed community-level change. To a large extent, however, the question of what factors determine the dynamics of an individual's pronunciation over time remains open, and this is the primary question raised for future work. Addressing this question through the study of medium-term pronunciation dynamics is key for understanding why accent change occurs (or does not occur) in individuals, the mechanism of sound change in communities, and the link between the two.

Appendix A: Static factors

This section describes the major static factors (covariates unrelated to time) affecting each phonetic variable, which are incorporated into the statistical models described in Appendix B of how each phonetic variable depends on static factors only. Table 2 in the main text summarizes the variables included in each static model, denoted in SMALL CAPS below.

We first define two static factors that are used as predictors for several phonetic variables. SPEECH RATE was calculated as the number of syllables per second (determined using the force-aligned transcriptions) within a phrase, defined as an interval of speech by a housemate bounded on each side by at least 60 ms of silence/nonspeech. We follow Kendall (2013) in using a rate measure that excludes pauses and in the choice of 60 ms as a threshold. WORD FREQUENCY was taken to be the token count per million of the orthographic word in the corpus of transcriptions of housemate speech, log-transformed.

A1. VOICE ONSET TIME. VOT is affected by a range of linguistic and social factors, reviewed in Auzou et al. 2000 and Docherty 1992.

Whether the stop consonant is (phonologically) voiced or voiceless is the single largest factor affecting VOT (voiced < voiceless). Because VOT behaves qualitatively differently for voiced and voiceless stops, it makes sense to think of these as two different variables, and we build separate statistical models for VOT of voiced and voiceless stops. VOT is strongly affected by PLACE OF ARTICULATION, with VOT expected to be progressively larger for bilabial, alveolar, and velar stops (e.g. Docherty 1992, Lisker & Abramson 1967, Nearey & Rochet 1994). VOT (in word-initial stops) is also influenced by the FOLLOWING SEGMENT IDENTITY and FOLLOWING VOWEL HEIGHT in the host word: VOT is expected to be higher before consonants (in complex onsets) than before vowels (in CV syllables), and higher before high vowels than before low vowels (e.g. Docherty 1992, Klatt 1975, Nearey & Rochet 1994). VOT is expected to be shorter for more frequent words (Stuart-Smith et al. 2015, Yao 2009, Yu et al. 2013), and to be greater in stressed or accented syllables, though this is more consistently found for voiceless stops than for phonologically voiced stops (Cole et al. 2007, Lisker & Abramson 1967, Stuart-Smith et al. 2015). VOT strongly decreases with increased speech rate for voiceless stops, but for voiced stops, VOT shows no correlation with speech rate or only a weak trend (e.g. Miller et al. 1986, Stuart-Smith et al. 2015). In the static models we consider both a speaker's mean speech rate (SPEECH RATE MEAN) and the deviation of speech rate from the speaker's mean for each token (SPEECH RATE DEVIATION), following Stuart-Smith and colleagues (2015).

VOT also differs between speakers: varieties of English, including British varieties (e.g. Docherty 1992, Docherty et al. 2011, Scobbie 2006), differ in the magnitude of positive VOT and the prevalence of prevoicing, and speakers of the same variety of English differ in their baseline VOT after controlling for differences in speech rate (Allen et al. 2003). SPEAKER GENDER may affect VOT (male < female; see Morris et al. 2008). By-speaker random-effect terms in our models account for interspeaker differences in VOT beyond speaker gender.

A2. CORONAL STOP DELETION. Deletion of word-final *t/d* in consonant clusters is affected by a variety of linguistic and social factors, reviewed by Hazen (2011), Schreier (2005), and Tagliamonte and Temple (2005).

CSD rate is affected by the context in which the host word (ending in *t/d*) is produced and properties of the host word. FOLLOWING CONTEXT strongly affects deletion rate, with deletion always found to be more frequent before consonants (C) than before vowels (V), and most frequent before another coronal stop (/t/ or /d/). Different studies find different effects of a following 'pause', which is usually treated as an alternative context (to C or V). Following Tanner and colleagues (2017), we instead consider PAUSE DURATION (following the *t/d*-site), acting as a proxy for prosodic boundary strength, as a variable that can affect CSD rate independently of following context; deletion is expected to be less likely for longer pauses. (Pause duration was manually transcribed during the annotation process.) PRECEDING PHONOLOGICAL CONTEXT affects CSD, with deletion generally found to be more likely after sonorants than after obstruents, and most likely after sibilants (especially /s/). CSD, as a reductive process, is expected to be more likely for higher-FREQUENCY words (e.g. Myers & Guy 1997, Walker 2012), at faster SPEECH RATES, and in more casual speech. MORPHOLOGICAL CLASS of the host word often affects CSD, with deletion progressively more likely in regular past-tense forms, irregular pasts, and nonpasts, though Tagliamonte and Temple (2005) notably failed to find any effect for York English.

While most social factors have weak effects on CSD rate, including speaker GENDER, different regional and ethnic varieties of English show very different deletion rates (Schreier 2005). We thus expect the speakers in our data set to differ greatly in overall CSD rate, due to their heterogeneous backgrounds, and we account for speaker differences beyond gender in the static model of CSD rate using by-speaker random-effect terms.

A3. VOWEL FORMANTS. Variation in F1 and F2 is conditioned by many properties of context, the utterance, and the speaker. First, vowel formants are strongly affected by coarticulation. The strongest coarticulatory effects come from the consonantal context: the manner, place, and voicing of adjacent consonants affect vowel formants in complex ways that are well documented for laboratory speech (e.g. Beddor 1982, Hillenbrand et al. 2001, Stevens & House 1963), as well as in sociolinguistic studies in (usually) spontaneous speech (Labov 1994), making standard sociolinguistic practice appropriate for our own (spontaneous speech) data set. Our models of F1 and F2 use a parametrization of consonantal context into four variables that is common in sociolinguistic studies: PRECEDING SEGMENT identity (ten levels), FOLLOWING SEGMENT MANNER of articulation (seven levels), FOLLOWING SEGMENT PLACE of articulation (seven levels), and FOLLOWING SEGMENT VOICING (three levels). We use the coding given by FAVE (Rosenfelder et al. 2011) for these variables, described in Table 2.

Prosody and speech style also strongly affect vowel formants, with vowels in more prominent positions or 'clearer' speech (indexed by slower speech rate, etc.) realized as less 'reduced' (e.g. more peripheral in formant space) relative to nonprominent positions or casual speech (e.g. Cho 2005, Fourakis 1991, Moon & Lindblom 1994). We considered both speech rate and vowel duration as correlates of clear speech and prominence. Only log-transformed VOWEL DURATION was included in the final models of F1 and F2, since vowel duration and speech rate were highly correlated and models including just vowel duration generally had a lower AIC than models including both terms or just speech rate. We excluded tokens from unstressed syllables (§3.3 'Annotation and data sets').

Formant frequencies for a given vowel vary greatly by speaker, both across dialect regions (see §3.3) and among speakers within a dialect region, due to physiological differences and social factors (e.g. Hillenbrand et al. 1995, Labov 2000). We do not explicitly model the effects of by-speaker factors on vowel formants, due to data sparsity, but account for speaker differences using by-speaker random-effect terms.

APPENDIX B: STATIC MODELS

This appendix describes the statistical models of how VOT and vowel formants depend on the static factors described above (Appendix A), as well as controlling for variability among speakers and words. Each static model is a mixed-effects regression model fitted using the lme4 package in R (Bates et al. 2014). The static model for CSD comes from Tanner et al. 2017, and relevant aspects are summarized in §4.2.

B1. VOICE ONSET TIME. Two linear mixed-effects models of log(VOT) were fit for voiced and voiceless stops for the VOT data set (§3.1 'Annotation and data set'). As fixed effects, each model contained main-effect terms for the nine static factors in Table 2 (top), as well as a term to account for annotator identity (ANNOTATOR), which was included as a factor with three levels for voiceless VOT and two levels for voiced VOT (the number of annotators for each subset). Each variable was coded such that the intercept of each model has the interpretation 'VOT, when all variables are held at their average values': factors are coded using Helmert contrasts, and continuous variables are STANDARDIZED by centering and dividing by two standard deviations (Gelman & Hill 2007). SPEECH RATE DEVIATION was further coded as a nonlinear spline with two components for the voiceless-stop data (using rcs in the rms package; Harrell 2014) and as a linear function for the voiced-stop data, based on exploratory plots of the relationship between this variable and log(VOT) for each subset. The models contained interaction terms as fixed effects, determined by exploratory data analysis (exploratory plots, stepwise model selection). The voiceless VOT model contained interactions between PLACE OF ARTICULATION and each of SPEECH RATE DEVIATION, STRESS, and FOLLOWING SEGMENT; the voiced VOT model contained interactions between PLACE OF ARTICULATION and each of SPEECH RATE DEVIATION, SPEECH RATE MEAN, and FOLLOWING SEGMENT. Both models contained all possible by-speaker and by-word random effects (Barr et al. 2013), with the exception of terms for annotator, but excluded correlations between random-effect terms, to aid convergence.

The results of the models are summarized in Table A1. Fixed-effect coefficients are shown with associated standard errors, degrees of freedom, test statistic (t), and significances, calculated using the Satterthwaite approximation using lmerTest (Kuznetsova et al. 2014). Random-effect terms are not shown.

B.2. VOWEL FORMANTS. Six linear mixed-effects regression models of formant realization—F1 and F2 for GOOSE, STRUT, and TRAP—were fitted for the vocalic data sets (§3.3 'Annotation and data sets').

Each model contained terms for each static factor in Table 2 (bottom): fixed-effect terms for VOWEL DURATION (log-transformed, then standardized) and FOLLOWING SEGMENT VOICING (coded using Helmert contrasts), and random intercept terms for FOLLOWING SEGMENT MANNER, FOLLOWING SEGMENT PLACE, and

Voiceless stops

COEFFICIENT	EST	*SE*	*df*	*t*	*P* (> *t*)
(intercept)	3.970	0.039	19.5	100.76	< 0.001
Speech rate deviation	−0.214	0.020	50.7	−10.58	< 0.001
Speech rate deviation′	0.034	0.009	33.1	3.91	< 0.001
Phrase position	0.071	0.017	16.9	4.08	< 0.001
Annotator	−0.014	0.006	12,061.8	−2.35	0.019
Annotator′	0.024	0.003	12,116.7	9.31	< 0.001
Place of articulation (P vs. T/K)	0.131	0.009	38.1	14.10	< 0.001
Place of articulation′ (T vs. K)	−0.116	0.019	33.8	−6.17	< 0.001
Stress	0.178	0.028	51.3	6.27	< 0.001
Following segment type	0.168	0.024	30.8	6.92	< 0.001
Vowel height	0.071	0.020	55.9	3.46	0.001
Frequency	−0.085	0.017	92.5	−5.10	< 0.001
Speech rate mean	−0.041	0.081	18.1	−0.51	0.617
Gender	−0.094	0.078	17.7	−1.21	0.244
POA : Speech rate deviation	0.023	0.014	69.5	1.59	0.116
POA′ : Speech rate deviation	−0.027	0.017	575.0	−1.61	0.108
POA : Speech rate deviation′	−0.007	0.006	3,209.1	−1.15	0.249
POA′ : Speech rate deviation′	0.007	0.009	160.6	0.82	0.411
POA : Stress	−0.057	0.020	34.4	−2.89	0.007
POA′ : Stress	0.103	0.033	102.1	3.17	0.002
POA : Following segment type	0.033	0.017	38.7	1.91	0.063
POA′ : Following segment type	−0.020	0.032	37.8	−0.62	0.540

Voiced stops

COEFFICIENT	EST	*SE*	*df*	*t*	*P* (> *t*)
(intercept)	2.766	0.050	18.4	55.62	< 0.001
Speech rate deviation	−0.047	0.017	20.5	−2.82	0.010
Phrase position	0.096	0.033	30.1	2.94	0.006
Annotator	0.128	0.006	12,777.4	22.10	< 0.001
Place of articulation (B vs. D/G)	0.225	0.011	26.1	20.07	< 0.001
Place of articulation′ (D vs. G)	0.003	0.028	22.1	0.11	0.910
Stress	−0.035	0.039	23.6	−0.91	0.370
Following segment type	0.522	0.040	30.5	13.21	< 0.001
Vowel height	0.080	0.026	37.1	3.15	0.003
Frequency	0.018	0.016	50.5	1.15	0.256
Speech rate mean	0.066	0.100	17.0	0.66	0.518
Gender	−0.150	0.072	22.0	−2.08	0.05
POA : Speech rate deviation	−0.015	0.009	11.6	−1.68	0.119
POA′ : Speech rate deviation	−0.029	0.022	20.9	−1.33	0.198
POA : Speech rate mean	−0.089	0.022	21.0	−4.03	< 0.001
POA′ : Speech rate mean	0.039	0.056	19.4	0.70	0.494
POA : Following segment type	0.035	0.023	28.5	1.54	0.135
POA′ : Following segment type	−0.236	0.050	16.9	−4.75	< 0.001

TABLE A1. Summary of fixed-effects coefficients in the static models of log(VOT) for voiceless (top) and voiced (bottom) stops. Primes indicate different contrasts for factors with multiple levels, or spline components (for SPEECH RATE DEVIATION).

PRECEDING SEGMENT. With this coding of static-factor variables, the intercept of each model has the interpretation of a grand mean.[18] Each model also contained by-speaker and by-item random intercepts (where item = instance of a vowel in a particular word, e.g. first TRAP vowel in *grandad*), a by-speaker random slope for VOWEL DURATION, and a term for its correlation with the by-speaker random intercept.

This model structure differs from that used for the VOT and CSD models in two ways, both of which are motivated by data sparsity, while keeping in mind the goal of the static models: to account for the effects of static factors on vowel formants so that we can then examine whether there is time dependence in housemates' pronunciations of vowels after accounting for these factors. First, some static factors are coded as random-effect terms, rather than as fixed effects. We do this for the static factors that have more than two levels

[18] Note that random intercept terms index deviations from the grand mean.

Fixed effects

	F1					F2				
GOOSE	EST	*SE*	*df*	*t*	*P* (> *t*)	EST	*SE*	*df*	*t*	*P* (> *t*)
(intercept)	−0.99	0.08	8.50	−13.21	0.00	0.00	0.31	10.97	−0.02	0.99
Vowel duration	0.02	0.03	11.93	0.77	0.46	−0.07	0.05	12.25	−1.38	0.19
Following C voicing	−0.04	0.07	3.81	−0.55	0.61	0.08	0.29	4.41	0.28	0.79
Following C	−0.01	0.03	4.20	−0.56	0.60	−0.03	0.10	4.68	−0.27	0.80
STRUT										
(intercept)	0.23	0.14	17.03	1.63	0.12	−1.09	0.14	23.01	−7.66	0.00
Vowel duration	0.35	0.05	11.60	7.38	0.00	−0.17	0.05	11.13	−3.15	0.01
Following C voicing	−0.12	0.09	8.45	−1.35	0.21	0.04	0.11	7.79	0.39	0.71
Following C	0.00	0.03	9.22	0.08	0.94	−0.01	0.04	8.27	−0.19	0.85
TRAP										
(intercept)	1.37	0.20	238.54	6.85	0.00	−0.38	0.16	101.45	−2.42	0.02
Vowel duration	0.50	0.05	11.22	10.48	0.00	0.02	0.04	11.45	0.44	0.67
Following C voicing	−0.02	0.28	413.74	−0.06	0.96	−0.05	0.19	284.74	−0.25	0.80
Following C	−0.06	0.09	418.17	−0.65	0.52	0.03	0.06	290.06	0.49	0.63

Random effects

	GOOSE		STRUT		TRAP	
	F1	F2	F1	F2	F1	F2
Preceding segment	0.07	0.39	0.079	0.184	0.065	0.136
Following segment manner	0.12	0.49	0.108	0.135	0.135	0.068
Following segment place	0.00	0.15	0.074	0.106	0.055	0.103

TABLE A2. Summary of static models of F1 and F2 for GOOSE, STRUT, and TRAP. Top: summary of fixed-effect coefficients. Primes indicate different contrasts for factors with multiple levels. Bottom: variance components (as standard deviations) for random-effect terms corresponding to properties of adjacent consonants.

(FOLLOWING SEGMENT MANNER, FOLLOWING SEGMENT PLACE, and PRECEDING SEGMENT), because these factors are not independent and the data are highly unbalanced among different levels of these factors,[19] making stable estimates of fixed-effect coefficients for these factors impossible. Second, there are no random slope terms accounting for differences among speakers in the effects of static factors with the exception of VOWEL DURATION, because the data are simply too sparse to estimate differences between speakers in the effect of phonological context, given the large number of contexts. The models thus assume that the effects of context are the same for all speakers. This assumption should not affect the findings of the dynamic models with respect to our research questions.

Table A2 summarizes the results of the models, with respect to the effects of static factors on F1 and F2 for the three vowels. Fixed-effect coefficients are shown with associated standard errors, degrees of freedom, test statistic (*t*), and significances, calculated as for VOT using lmerTest. Also shown are the variance components for each random intercept corresponding to a static factor. Other random-effect terms are not shown.

REFERENCES

AGRESTI, ALAN. 2002. *Categorical data analysis*. 2nd edn. Hoboken, NJ: Wiley.

ALLEN, J. SEAN; JOANNE L. MILLER; and DAVID DESTENO. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustic Society of America* 113.544–52. DOI: 10.1121/1.1528172.

AUER, PETER; BIRGIT BARDEN; and BEATE GROßKOPF. 1998. Subjective and objective parameters determining 'salience' in long-term dialect accommodation. *Journal of Sociolinguistics* 2.163–87. DOI: 10.1111/1467-9481.00039.

AUER, PETER, and FRANS HINSKENS. 2005. The role of interpersonal accommodation in a theory of language change. In Auer et al., 335–57.

AUER, PETER; FRANS HINSKENS; and PAUL KERSWILL (eds.) 2005. *Dialect change: Convergence and divergence in European languages*. Cambridge: Cambridge University Press.

[19] For example: there is no bilabial lateral in English, and among the 5,365 TRAP tokens, only four are before a dental consonant.

Auzou, Pascal; Canan Ozsancak; Richard J. Morris; Mary Jan; Francis Eustache; and Didier Hannequin. 2000. Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics* 14.131–50. DOI: 10.1080/02699 2000298878.

Baayen, R. Harald; Douglas J. Davidson; and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412. DOI: 10.1016/j.jml.2007.12.005.

Baayen, R. Harald; Richard Piepenbrock; and Leon Gulickers. 1996. *CELEX2* [CD-ROM]. LDC96L14. Philadelphia: Linguistic Data Consortium.

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39.437–56. DOI: 10.1017/S0047404510000400.

Babel, Molly. 2011. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40.177–89. DOI: 10.1016/j.wocn.2011.09.001.

Babel, Molly; Andrew Garrett; Michael J. Houser; and Maziar Toosarvandani. 2013. Descent and diffusion in language diversification: A study of Western Numic dialectology. *International Journal of American Linguistics* 79.445–89. DOI: 10.1086/67 1772.

Baker, Adam; Diana Archangeli; and Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23.347–74. DOI: 10.1017/S0954394511000135.

Bane, Max; Peter Graff; and Morgan Sonderegger. 2010. Longitudinal phonetic variation in a closed system. *Chicago Linguistic Society* 46(1).43–58.

Barden, Birgit, and Beate Großkopf. 1998. *Sprachliche Akkommodation und soziale Integration* [Speech accommodation and social integration]. Tübingen: Niemeyer.

Barr, Dale J.; Roger Levy; Christoph Scheepers; and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–78. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas; Martin Maechler; Ben Bolker; and Steven Walker. 2014. lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-7. Online: https://cran.r-project.org/web/packages/lme4/index.html.

Beal, Joan. 2004. English dialects in the North of England: Phonology. In Schneider et al., 113–33.

Beddor, Patrice S. 1982. *Phonological and phonetic effects of nasalization on vowel height*. Minneapolis: University of Minnesota dissertation.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13.145–204. DOI: 10.1017/S004740450001037X.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Boersma, Paul, and David Weenink. 2011. Praat: Doing phonetics by computer. Version 5.2.17. Online: http://www.praat.org/.

Bourhis, Richard Y., and Howard Giles. 1977. The language of intergroup distinctiveness. *Language, ethnicity, and intergroup relations*, ed. by Howard Giles, 119–35. London: Academic Press.

Bowie, David. 2005. Language change over the lifespan: A test of the apparent time construct. *University of Pennsylvania Working Papers in Linguistics* 11.45–58. Online: http://repository.upenn.edu/pwpl/vol11/iss2/3.

Breen, Mara; Evelina Fedorenko; Michael Wagner; and Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25. 1044–98. DOI: 10.1080/01690965.2010.504378.

Brink, Lars, and Jørn Lund. 1975. *Dansk rigsmål: Lydudviklingen siden 1840 med særligt henblik på sociolekterne i københavn*. Copenhagen: Gyldendal.

Chambers, Jack K. 2003. *Sociolinguistic theory: Linguistic variation and its social significance*. 2nd edn. Malden, MA: Blackwell.

Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *The Journal of the Acoustical Society of America* 117.3867–78. DOI: 10.1121/1.1861893.

Chodroff, Eleanor; John Godfrey; Sanjeev Khudanpur; and Colin Wilson. 2015. Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org /icphs-proceedings/ICPhS2015/Papers/ICPHS0632.pdf.

Clark, Urszula. 2004. The English West Midlands: Phonology. In Schneider et al., 134–62.
Cole, Jennifer; Heejin Kim; Hansook Choi; and Mark Hasegawa-Johnson. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics* 35.180–209. DOI: 10.1016/j.wocn.2006.03.004.
Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9.1–12. DOI: 10.1017/S0047404500007752.
Cox, Felicity, and Sallyanne Palethorpe. 2007. Australian English. *Journal of the International Phonetic Association* 37.341–50. DOI: 10.1017/S0025100307003192.
Cukor-Avila, Patricia. 2015. Growing INTO and IN African American Vernacular English. Paper presented at the 10th UK Language Variation and Change (UKLVC) conference, University of York.
Cukor-Avila, Patricia, and Guy Bailey. 2013. Real time and apparent time. *The handbook of language variation and change*, 2nd edn., ed. by Jack K. Chambers and Natalie Schilling, 237–62. Malden, MA: Wiley-Blackwell.
Darlington, Richard B., and Tom V. Smulders. 2001. Problems with residual analysis. *Animal Behaviour* 62.599–602. DOI: 10.1006/anbe.2001.1806.
Delvaux, Véronique, and Alain Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64.145–73. DOI: 10.1159/000107914.
Docherty, Gerard. 1992. *The timing of voicing in British English obstruents*. Berlin: Foris.
Docherty, Gerard. 2010. Phonological innovation in contemporary spoken British English. *The Routledge handbook of world Englishes*, ed. by Andy Kirkpatrick, 59–75. London: Routledge.
Docherty, Gerard; Dominic Watt; Carmen Llamas; Damien Hall; and Jennifer Nycz. 2011. Variation in voice onset time along the Scottish-English border. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 591–94. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Docherty/Docherty.pdf.
Eckert, Penelope. 2000. *Linguistic variation as social practice*. Malden, MA: Blackwell.
Evans, Bronwen G., and Paul Iverson. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America* 121.3814–26. DOI: 10.1121/1.2722209.
Ferragne, Emmanuel, and François Pellegrino. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association* 40.1–34. DOI: 10.1017/S0025100309990247.
Ferreira, Fernanda, and Charles Clifton, Jr. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25.348–68. DOI: 10.1016/0749-596X(86)90006-9.
Fine, Alex B.; T. Florian Jaeger; Thomas A. Farmer; and Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8:e77661. DOI: 10.1371/journal.pone.0077661.
Foreman, Annik. 2003. *Pretending to be someone you're not: A study of second dialect acquisition in Australia*. Melbourne: Monash University dissertation.
Fourakis, Marios. 1991. Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America* 90.1816–27. DOI: 10.1121/1.401662.
Fowler, Carol A.; Julie M. Brown; Laura Sabadini; and Jeffrey Weihing. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* 49.396–413. DOI: 10.1016/S0749-596X(03)00072-X.
Freckleton, Robert P. 2002. On the misuse of residuals in ecology: Regression of residuals vs. multiple regression. *Journal of Animal Ecology* 71.542–45. DOI: 10.1046/j.1365-2656.2002.00618.x.
Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press.
Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Giles, Howard; Justine Coupland; and Nikolas Coupland. 1991. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, ed. by Howard Giles, Justine Coupland, and Nikolas Coupland, 1–68. Cambridge: Cambridge University Press.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.251–79. DOI: 10.1037/0033-295X.105.2.251.

Goldinger, Stephen D. 2000. The role of perceptual episodes in lexical processing. *Proceedings of the workshop on spoken word access processes*, ed. by Anne Cutler, James M. McQueen, and Rian Zondervan, 155–58. Nijmegen: Max Planck Institute.

Good, Phillip. 2013. *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.

Gregersen, Frans; Torben Juel Jensen; and Nicolai Pharao. 2017. Comparing speech samples: On the challenge of comparability in panel studies of language change in real time. *Panel studies of variation and change*, ed. by Suzanne Evans Wagner and Isabelle Buchstaller. London: Routledge, to appear.

Gregersen, Frans; J. Normann Jørgensen; Janus Spindler Møller; Nicolai Pharao; and Gert Foget Hansen. 2017. Sideways: Five methodological studies of sociolinguistic interviews. *Acta Linguistica Hafniensia*. DOI: 10.1080/03740463.2017.1335111.

Haddican, Bill; Paul Foulkes; Vincent Hughes; and Hazel Richards. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change* 25.371–403. DOI: 10.1017/S0954394513000197.

Harrell, Frank E., Jr. 2014. rms: Regression modeling strategies. R package version 4.2-0. Online: http://CRAN.R-project.org/package=rms.

Harrington, Jonathan; Sallyanne Palethorpe; and Catherine Watson. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30.63–78. DOI: 10.1017/S0025100300006666.

Hastie, Trevor J., and Robert J. Tibshirani. 1990. *Generalized additive models*. Boca Raton, FL: CRC.

Hawkins, Sarah, and Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35.183–99. DOI: 10.1017/S0025100305002124.

Hazen, Kirk. 2011. Flying high above the social radar: Coronal stop deletion in modern Appalachia. *Language Variation and Change* 23.105–37. DOI: 10.1017/S0954394510000220.

Heald, Shannon L. M. 2012. *Tuning the targets of speech production: Effects of articulatory and perceptual experience on speech production*. Chicago: University of Chicago dissertation.

Hillenbrand, James; Michael J. Clark; and Terrance M. Nearey. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America* 109.748–63. DOI: 10.1121/1.1337959.

Hillenbrand, James; Laura A. Getty; Michael J. Clark; and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97.3099–3111. DOI: 10.1121/1.411872.

Hindle, Donald. 1980. *The social and structural conditioning of phonetic variation*. Philadelphia: University of Pennsylvania dissertation.

Hockett, Charles F. 1965. Sound change. *Language* 41.185–204. DOI: 10.2307/411873.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145–65. San Diego: Academic Press.

Katseff, Shira, and John Houde. 2008. Partial compensation in speech adaptation. *UC Berkeley Phonology Lab Annual Report (2008)*, 444–61. Online: http://escholarship.org/uc/item/2rm0853c.

Katseff, Shira; John Houde; and Keith Johnson. 2012. Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech* 55.295–308. DOI: 10.1177/0023830911417802.

Kendall, Tyler. 2013. *Speech rate, pause and sociolinguistic variation*. Basingstoke: Palgrave Macmillan.

Kerswill, Paul, and Ann Williams. 2002. 'Salience' as an explanatory factor in language change: Evidence from dialect levelling in urban England. *Language change: The interplay of internal, external and extra-linguistic factors*, ed. by Mari C. Jones and Edith Esch, 81–110. Berlin: Mouton de Gruyter.

Keshet, Joseph; Morgan Sonderegger; and Thea Knowles. 2014. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction. Version 0.91. Online: https://github.com/mlml/autovot/.

Kim, Midam; William S. Horton; and Ann R. Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2.125–56. DOI: 10.1515/labphon.2011.004.

Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research* 18.686–706. DOI: 10.1044/jshr.1804.686.

Kryuchkova, Tatiana; Benjamin V. Tucker; Lee H. Wurm; and R. Harald Baayen. 2012. Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language* 122.81–91. DOI: 10.1016/j.bandl.2012.05.005.

Kuznetsova, Alexandra; Per Bruun Brockhoff; and Rune Haubo Bojesen Christensen. 2014. lmerTest: Tests in linear mixed effects models. R package version 2.0-11. Online: http://CRAN.R-project.org/package=lmerTest.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2.205–54. DOI: 10.1017/S0954394500000338.

Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2000. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell.

Labov, William. 2007. Transmission and diffusion. *Language* 83.344–87. DOI: 10.1353/lan.2007.0082.

Le Page, Robert Brock, and Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.

Lisker, Leigh, and Arthur S. Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10.1–28. DOI: 10.1177/002383096701000101.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49.606–8. DOI: 10.1121/1.1912396.

MacLeod, Bethany. 2012. *The effect of perceptual salience on cross-dialectal phonetic convergence in Spanish*. Toronto: University of Toronto dissertation.

Miller, Joanne L.; Kerry P. Green; and Adam Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43.106–15. DOI: 10.1159/000261764.

Milroy, James, and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21.339–84. DOI: 10.1017/S0022226700010306.

Mitterer, Holger, and Mirjam Ernestus. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109.168–73. DOI: 10.1016/j.cognition.2008.08.002.

Moon, Seung-Jae, and Björn Lindblom. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America* 96.40–55. DOI: 10.1121/1.410492.

Morris, Richard J.; Christopher R. McCrea; and Kaileen D. Herring. 2008. Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics* 36.308–17. DOI: 10.1016/j.wocn.2007.06.003.

Myers, James, and Gregory R. Guy. 1997. Frequency effects in variable lexical phonology. *University of Pennsylvania Working Papers in Linguistics* 4.215–27. Online: http://repository.upenn.edu/pwpl/vol4/iss1/14.

Nahkola, Kari, and Marja Saanilahti. 2004. Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change* 16.75–92. DOI: 10.1017/S0954394504162017.

Namy, Laura L.; Lynne C. Nygaard; and Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21.422–32. DOI: 10.1177/026192702237958.

Nearey, Terrance M., and Bernard L. Rochet. 1994. Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association* 24.1–19. DOI: 10.1017/S002510030000 4965.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39.132–42. DOI: 10.1016/j.wocn.2010.12.007.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119.2382–93. DOI: 10.1121/1.2178 720.

Pardo, Jennifer S. 2009. Expressing oneself in conversational interaction. *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, ed. by Ezequiel Morsella, 183–96. Mahwah, NJ: Lawrence Erlbaum.

Pardo, Jennifer S. 2013. Measuring phonetic convergence in speech production. *Frontiers in Psychology* 4:559. DOI: 10.3389/fpsyg.2013.00559.

Pardo, Jennifer S.; Rachel Gibbons; Alexandra Suppes; and Robert M. Krauss. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40.190–97. DOI: 10.1016/j.wocn.2011.10.001.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. 1st edn. Halle: Max Niemeyer.

Penhallurick, Robert. 2004. Welsh English: Phonology. In Schneider et al., 98–112.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan L. Bybee and Paul J. Hopper, 137–58. Amsterdam: John Benjamins.

Pinheiro, José, and Douglas Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pisoni, David B. 1980. Variability of vowel formant frequencies and the quantal theory of speech: A first report. *Phonetica* 37.285–305. DOI: 10.1159/000259999.

Prince, Ellen F. 1987. Sarah Gorby, Yiddish folksinger: A case study of dialect shift. *International Journal of the Sociology of Language* 67.83–116. DOI: 10.1515/ijsl.1987 .67.83.

R Core Team. 2014. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org/.

Rickford, John R., and Faye McNair-Knox. 1994. Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. *Sociolinguistic perspectives on register*, ed. by Douglas Biber and Edward Finegan, 235–76. Oxford: Oxford University Press.

Rickford, John R., and Mackenzie Price. 2013. Girlz II women: Age-grading, language change, and stylistic variation. *Journal of Sociolinguistics* 17.143–79. DOI: 10.1111 /josl.12017.

Rosenfelder, Ingrid; Josef Fruehwald; Keelan Evanini; and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction). Online: http://fave.ling.upenn.edu.

Sancier, Michele L., and Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25.421–36. DOI: 10.1006/jpho .1997.0051.

Sankoff, Gillian. 2005. Cross-sectional and longitudinal studies in sociolinguistics. *Sociolinguistics: An international handbook of the science of language and society*, ed. by Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, and Peter Trudgill, vol. 2, 1003–13. Berlin: De Gruyter.

Sankoff, Gillian. 2013. Longitudinal studies. *The Oxford handbook of sociolinguistics*, ed. by Robert Bayley, Richard Cameron, and Ceil Lucas, 261–79. Oxford: Oxford University Press.

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.560–88. DOI: 10.1353/lan.2007.0106.

Schneider, Edgar W.; Kate Burridge; Bernd Kortmann; Rajend Mesthrie; and Clive Upton (eds.) 2004. *A handbook of varieties of English, vol. 1: Phonology*. New York: Mouton de Gruyter.

Schreier, Daniel. 2005. *Consonant change in English worldwide*. New York: Palgrave Macmillan.

Schwartz, Misha. 2015. Formant visualization and remeasurement (FVR). Online: https://github.com/mlml/fvr.

Scobbie, James M. 2006. Flexibility in the face of incompatible English VOT systems. *Laboratory phonology 8*, ed. by Louis M. Goldstein, Doug H. Whalen, and Catherine T. Best, 367–92. Berlin: Mouton de Gruyter.

Siegel, Jeff. 2010. *Second dialect acquisition*. Cambridge: Cambridge University Press.

Sonderegger, Morgan. 2012. *Phonetic and phonological dynamics on reality television*. Chicago: University of Chicago dissertation.

Sonderegger, Morgan. 2015. Trajectories of voice onset time in spontaneous speech on reality TV. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0903.pdf.

Sonderegger, Morgan, and Joseph Keshet. 2012. Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America* 132.3965–79. DOI: 10.1121/1.4763995.

Stanford, James N. 2008. A sociotonetic analysis of Sui dialect contact. *Language Variation and Change* 20.409–50. DOI: 10.1017/S0954394508000161.

Stevens, Kenneth N., and Arthur S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6.111–28. DOI: 10.1044/jshr.0602.111.

Stuart-Smith, Jane. 2004. Scottish English: Phonology. In Schneider et al., 47–67.

Stuart-Smith, Jane; Morgan Sonderegger; Tamara Rathcke; and Rachel MacDonald. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology* 6.505–49. DOI: 10.1515/lp-2015-0015.

Tagliamonte, Sali, and Rosalind Temple. 2005. New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17.281–302. DOI: 10.1017/S0954394505050118.

Tanner, James; Morgan Sonderegger; and Michael Wagner. 2017. Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology* 8(1):15. DOI: 10.5334/labphon.96.

Temple, Rosalind A. M. 2009. (t,d): The variable status of a variable rule. *Oxford University Working Papers in Linguistics, Philology & Phonetics* 12.145–70.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Oxford: Oxford University Press.

Tsukada, Kimiko. 2008. An acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers. *English World-Wide* 29.194–211. DOI: 10.1075/eww.29.2.05tsu.

van Rij, Jacolien; Martijn Wieling; R. Harald Baayen; and Hedderik van Rijn. 2016. itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.2. Online: https://cran.r-project.org/web/packages/itsadug/.

Walker, James. 2012. Form, function, and frequency in phonological variation. *Language Variation and Change* 24.397–415. DOI: 10.1017/S0954394512000142.

Watson, Kevin. 2007. Liverpool English. *Journal of the International Phonetic Association* 37.351–60. DOI: 10.1017/S0025100307003180.

Wells, John C. 1982. *Accents of English*. 3 vols. Cambridge: Cambridge University Press.

Wieling, Martijn; Simonetta Montemagni; John Nerbonne; and R. Harald Baayen. 2014. Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language* 90.669–92. DOI: 10.1353/lan.2014.0064.

Wikipedia. 2012a. Big Brother 9 (UK). *Wikipedia, The Free Encyclopedia*. Online: http://en.wikipedia.org/wiki/Big_Brother_2008_(UK), accessed 21 April 2012.

Wikipedia. 2012b. Big Brother (UK). *Wikipedia, The Free Encyclopedia*. Online: http://en.wikipedia.org/wiki/Big_Brother_(UK), accessed 21 April 2012.

Wilson, James. 2010. *Moravians in Prague: A sociolinguistic study of dialect contact in the Czech Republic*. Frankfurt: Peter Lang.

Winter, Bodo, and Martijn Wieling. 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution* 1.7–18. DOI: 10.1093/jole/lzv003.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall/CRC.

Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B* 73.3–36. DOI: 10.1111/j.1467-9868.2010.00749.x.

Wood, Simon N. 2015. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. R package version 1.8–7. Online: https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/mgcv-package.html.

Yao, Yao. 2009. Understanding VOT variation in spontaneous speech. *Current numbers in unity and diversity of languages*, ed. by Manghyu Pak, 1122–37. Seoul: Linguistic Society of Korea.

Young, Steve; Gunnar Evermann; Mark Gales; Thomas Hain; Dan Kershaw; Xunying Liu; Gareth Moore; Julian Odell; Dave Ollason; Dan Povey; et al. 2006. *The HTK book (for HTK version 3.4)*. Cambridge: Cambridge University Engineering Department.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 201–27. Oxford: Oxford University Press.

Yu, Alan C. L.; Carissa Abrego-Collier; and Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits. *PLoS ONE* 8:e74746. DOI: 10.1371/journal.pone.0074746.

[morgan.sonderegger@mcgill.ca]
[max@thebanes.org]
[graffmail@gmail.com]

[Received 26 September 2015;
revision invited 18 April 2016;
revision received 4 August 2016;
accepted 16 October 2016]