Title
Accommodation Can Lead to Innovated Variation

Journal
UC Berkeley PhonLab Annual Report, 10(10)

ISSN
2768-5047

Author
Heath, Jevon

Publication Date
2014

DOI
10.5070/P78f86j4k3

Accommodation can lead to innovated variation

Jevon Heath
University of California, Berkeley

1. Introduction

The initiation of sound change requires the emergence of innovated variants. One possible source of innovated variants is non-faithful[1] transference of linguistic features during accommodation to a received speech signal. If the subphonemic cues associated with a phonological contrast are realized differently in an accommodating speaker's speech than in both the speaker's previous speech and in the signal to which they are accommodating, the resulting difference in phonetic realization may constitute a new variation.

In this paper I argue that this sort of difference logically must happen in accommodation, at least for some speakers along some phonetic dimensions. I then report the results of a study demonstrating that such phonetic divergence in accommodation does happen in a laboratory environment, indicating that accommodation is a potential source of new phonetic variants. I conclude by situating these findings within extant work on sound change. To the extent that populations of individuals evince varying degrees of non-faithful transference of these cues, this study provides evidence for a potential approach to the actuation problem (Weinreich et al. 1968).

2. Background

2.1 Defining accommodation

Accommodation is the phenomenon in which a person talks differently in response to external phonetic exposure. An often-reported example is the propensity of people calling their parents to "revert" to their speech. While "talking differently" is most commonly understood as talking more similarly to the received phonetic material (convergence), talking less similarly (divergence) has also been attested (cf. Giles 1971, Bourhis & Giles 1977, Babel 2010). However, divergence has only been evinced as moderated or intentional alteration of one's speech (cf. Bourhis & Giles 1977, Babel 2010), whereas convergence occurs both consciously and unconsciously (Bourhis & Giles 1977; Goldinger 1998). As such, convergence has commonly been considered default accommodative linguistic behavior.

[1]By *non-faithful transference* I mean a lack of phonetic fidelity between the signal perceived and the signal produced. I do not mean to invoke the idea of phonological faithfulness.

The external phonetic exposure acting as the catalyst for accommodation can be from a live individual (cf. Pardo 2006, Pardo et al. 2010), or a recorded speaker with manipulated or natural speech (cf. Bourhis & Giles 1977, Goldinger 1998), which may or may not be directed at the listener (Delvaux & Soquet 2007); accommodation occurs in each of these contexts. Put broadly, accommodation can include effects such as syntactic priming (Bock 1986) and word choice. However, it is intrinsically difficult to demonstrate that syntactic and semantic changes in behavior are directly attributable to an interlocutor's speech patterns rather than to other contextual factors[2]. Additionally, given that speakers exhibit variation all the time, a systematic pattern of change needs to be observed in order for differences to be chalked up to accommodation rather than to random variation. For these reasons studies of accommodation are often restricted to differences at the phonetic level. Previous studies have identified phonetic convergence in subphonemic features including lip movement (Gentilucci & Bernardis 2007), segment durations (Delvaux & Soquet 2007), voice onset time (Fowler et al. 2003, Shockley et al. 2004), and vowel formant frequency (Babel 2010); and discourse-level features including subvocal frequency/amplitude contour and fundamental frequency (F0) (Gregory 1990), pause and utterance duration (Jaffe & Feldstein 1970), speech rate (Giles et al. 1991), and vocal intensity (Natale 1975)[3]. Conversely, relatively few studies investigating convergence have reported phonetic features that do not exhibit it to some degree. Mitterer & Ernestus (2008) showed that on a by-trial basis, Dutch speakers imitated the presence of pre-voicing in producing initial voiced stops in nonce words, but did not imitate the amount of pre-voicing. This was attributed to the lack of phonological information encoded in the feature under investigation, as the duration of pre-voicing is not phonologically relevant in Dutch. Nielsen (2011) showed that VOT shortening is not accommodated in English voiceless stops in laboratory conditions, attributing this to an avoidance of the phonological ambiguity with voiced stops, generally realized in English as short-lag voiceless stops, that would result.

Phonetic convergence toward an interlocutor's speech begins within minutes of the initiation of discourse (Goldinger 1998; Nye & Fowler 2003; Pardo et al. 2010). In a perceptual similarity task, Goldinger & Azuma (2004) demonstrated that speech produced a week after hearing recorded stimuli in a laboratory environment sounded more similar to those stimuli than speech produced a day before hearing them, indicating that accommodation effects can persist up to at least a week after the initial interaction. The immediacy and persistence of its effects suggest that accommodation may be a natural

---

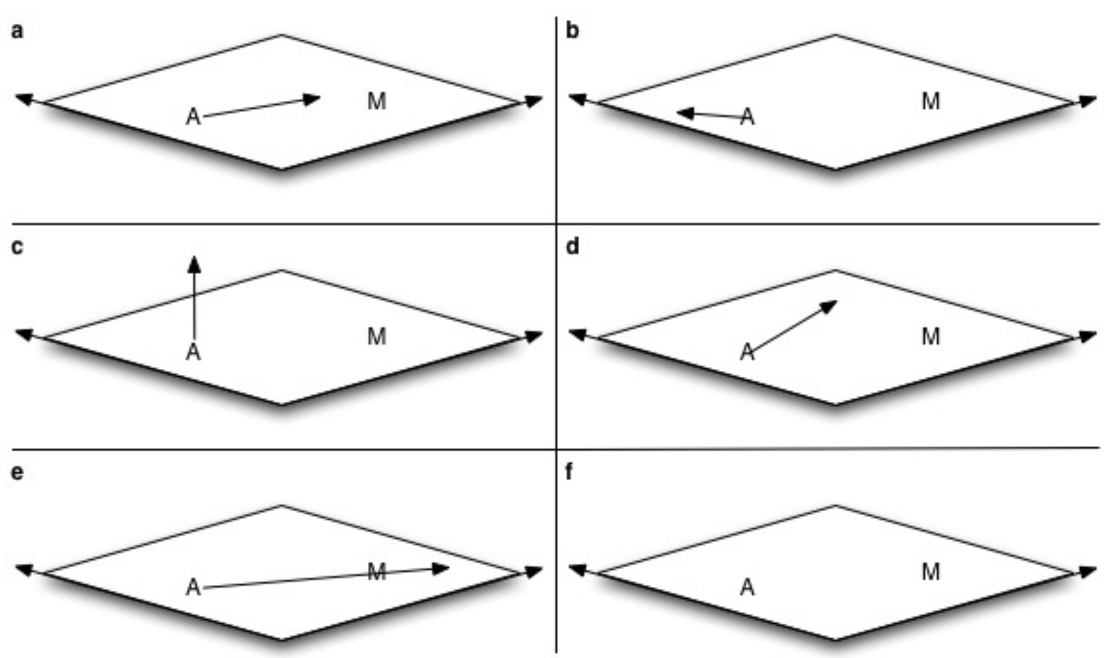
[2]Throughout this paper, I use *interlocutor* as a term of convenience to cover all forms of received phonetic exposure. This includes recorded and/or manipulated speech, as well as heard speech not directed at the listener.
[3]All of these studies dealt with varieties of English except for Delvaux & Soquet (2007), which examined Belgian French.

source of variation leading to permanent changes in individuals' speech patterns. Delvaux & Soquet (2007) hypothesize that imitation in speech explains both the stability of phonetic realizations within a speech community, and the potential for change to a community's speech norms via the transmission of new realizations. However, it may be possible to impute not only the transmission of sound change, but also the *actuation* of sound change, to the effects of accommodation. In other words, can accommodation alone – in the abstract, without appeal to supplemental considerations – directly result in the introduction of innovative forms? The wide-ranging nature and robustness of accommodation effects suggests that this idea is worth considering.

2.2 Directions of accommodation

As mentioned above, *convergence* and *divergence* are the two types of adjustment that have been discussed in literature on accommodation, along with *null accommodation* (a lack of adjustment). However, previous studies have generally looked at only one linguistic feature at a time, whereas multiple features may be manipulated at once in the course of speech production. When taking multiple dimensions into account at once, there are a total of four logically possible directions for accommodation, two of which have heretofore not been investigated. These directions are schematized in Figure 1.

Figure 1: Types of Accommodation

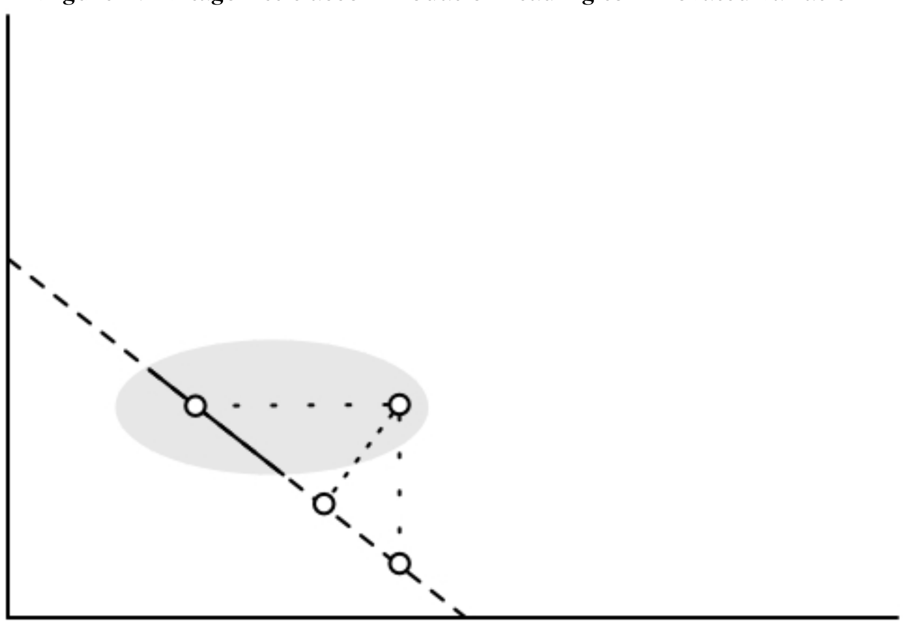


a. Convergence. b. Divergence. c. Orthogonal accommodation. d. Antagonistic accommodation. e. Hyperconvergence. f. Null accommodation. A = Accommodator, M = Model.

The third logically possible direction of accommodation is talking *orthogonally* like the other person, i.e. changing one's own speech in response to an interlocutor's speech, but in a manner not reflected in that interlocutor's speech. An example would be speaking in a whisper in response to a statement about a sleeping baby. Orthogonal accommodation to information on a phonetic level has not been attested, but nor has it been investigated. Should it exist, phonetic orthogonal accommodation would most likely be related to the indexicality of linguistic features. As its occurrence would be essentially impossible to isolate, I will henceforth ignore orthogonal accommodation.

The fourth logical possibility involves talking more like the other person in some ways and less like the other person in other ways. I am calling this possibility *antagonistic* accommodation, in the sense that the cues that are variously convergent and divergent are acting against each other. Antagonistic accommodation is a likely domain within which sound change actuation may happen, via the recoupling of gestures in new ways due to the pressures placed on an individual's speech system by either physiological or psychological/phonological factors. Figure 2 shows a schematization of this potentiality.

Figure 2: Antagonistic accommodation leading to innovated variation



The *x* and *y*-axes represent two cues to a phonetic feature that are in a physiologically or psychologically-induced antagonistic relationship: within an individual's speech, *x* increases as *y* decreases, and vice versa. The shaded area is the extant area of variation within the speech community for the feature in question. The dotted line represents the (physiologically or phonologically) restricted range of variation the speaker has available for the two cues in question; the solid line represents the speaker's available range within extant variation. The point not on the dashed line represents received speech from a model/interlocutor; the dotted lines represent the shortest distances between the received speech and the speaker's available range of variation, depending on whether one, the other, or both cues are adjusted.

When the speaker adjusts only the cue represented by the *x*-axis in convergence to the model, the resulting token is outside of the community's extant variation; when the speaker adjusts only the *y*-axis cue, the result is within the established range of variation.[4] However, in the event that a speaker adjusts both cues in order to approximate the model's variant as closely as possible along both axes, the resulting production is also outside of the community's extant variation. In converging to a greater degree along the *x*-axis cue, the speaker is diverging from the model along the *y*-axis cue.

2.3 Non-faithful accommodation

Unless accommodation can result in non-faithful transference of linguistic features from the received signal, it cannot explain the appearance of innovative forms within a population. However, if it *can* – if accommodation can result in the appearance of entirely new linguistic material – then accommodation is a possible source of permanent changes to the speech patterns of whole communities.

Previous studies have shown that speakers do evince non-faithful accommodation. Nielsen (2011) looked at whether accommodation can lead to a phonological generalization of received subphonemic features. Participants were exposed to model speech with extended VOTs exclusively in words beginning with /p/, and then recorded saying words beginning with both /p/ and /k/. Nielsen found that VOT was indeed generalized to /k/ in this fashion, although new words with initial /p/ evinced a higher degree of VOT adjustment than words with initial /k/. Additionally, words with a lower lexical frequency showed a higher degree of VOT adjustment than more frequent words. Nielsen's study established that non-faithful transference is possible in accommodation. However, the particular type of transference investigated by Nielsen does not result in a new pattern of phonetic material, as /k/ with extended VOT is found in English in general, and there was no VOT increase evinced in segments without such attested extensions, for instance /m/.

As of yet it has not been demonstrated whether accommodation can result in changes that were not present at all in the speech signal being accommodated toward. If it cannot – if all accommodation results in faithful changes to the received signal – the

[4]The comparatively narrow range of existing variation along the *y*-axis is aline with a feature that is relatively salient to speakers, whereas the wider range of variation along the *x*-axis suggests that the target in production is not so specific.

process can only result in the mingling of already extant linguistic features in the speech profiles of disparate individuals within a speech community or population. Similarly, changes that differ only in degree will not affect the constitution of a category unless those changes result in the extension of the category's boundaries. If a category has two non-overlapping areas in phonetic space associated with it, it is possible for a change in degree to lie between those areas. But for a contiguous and convex category, adjustment toward an already-extant variant will necessarily lie within the category as previously defined. Adjustment *away from* an already-extant variant, on the other hand, may well extend category boundaries.

Antagonistic accommodation is not the only type of accommodative behavior that is a potential source of innovative material. While convergence toward the interlocutor will not result in the introduction of new variations into the language, *hyperconvergence* – accommodation toward *and past* the speech of the interlocutor – may result in new variation. To the extent that they are attributable to automatic (*i.e.*, physiologically determined) imitation (Gentilucci & Bernardis 2007), both antagonistic accommodation and hyperconvergence provide speaker-level sources of community-level innovation, referred to by Chang (2012) as "phonetic drift" (repurposed from Sapir 1921). A change stemming from divergence could also conceivably result in new variation, to wit the extension of a category away from the examples provided by an interlocutor. As mentioned, however, it has not been demonstrated that such changes persist beyond the interaction in which they were instantiated. Given this and the restrictions of context and situation on divergence, it is perhaps unlikely that changes due to divergence would be picked up as new utile variants.

Of course, it is not necessarily the case that different individuals will necessarily evince the same type of accommodative behavior in the same situation. Differences in accommodation may stem from personal factors including but not limited to: prior experience with the linguistic features being encountered; intensity of desire to establish social acceptance or identity, in the context of the given interlocution as well as more generally; physiological or neurological idiosyncrasies directly impinging on speech perception and/or production; fatigue; inattention; and emotional state. Some of these factors are potentially useful in predicting accommodative behavior. Researchers have investigated the effects of factors such as social identification (Bourhis & Giles 1977), liking (Babel 2012), power (Pardo et al. 2010), and empathy (Abrego-Collier et al. 2011) on whether or not, and to what extent, speakers display accommodation effects.

3 The current study

The current study addresses the question of whether antagonistic accommodation happens by looking at *coincident cues*: features that tend to coincide with a given linguistic feature, signaling or reinforcing the identity of that feature. This type of reinforcement of information-laden units is exceedingly common on all levels of linguistic structure. In this paper, I will look at phonetic reinforcement, namely coincident cues to lengthened voice onset time (VOT) in voiceless stops in English. The study discussed here examines the degree of automaticity in accommodation to three cues associated with VOT – duration of closure of the stop in question, as well as both the initial F0 and total duration of the vowel following the stop – when only the VOT is experimentally manipulated.

3.1 Coincident cues of VOT

The current experiment looks at coincident cues of VOT for several reasons. VOT is generally held to be the most information-laden feature marking the phonological distinction between voiced and voiceless stops in English, as phonologically voiced stops are typically realized as voiceless, especially in onsets of stressed syllables. Previous studies have shown that the length of VOT of English voiceless stops can be manipulated in phonetic accommodation (Fowler et al. 2003; Shockley et al. 2004). VOT is relatively easy to measure with consistency, due to the clearly definable features signifying its onset (stop release) and endpoint (glottal pulse). Finally, there are many coincident cues of VOT in English with varying degrees of automaticity, three of which are under investigation in this study.

Stop closure duration: VOT is inversely correlated with stop closure duration (Lehiste 1970, Boucher 2002). For a given speaker, overall timing relations are such that the overall stop duration will remain approximately constant: as such, an increase in VOT is concurrent with a decrease in closure duration, and vice versa.

Initial F0 of following vowel: Higher onset vowel F0 is a perceptual cue for voiceless stops in English (House & Fairbanks 1953). However, there is no direct correlation between VOT and F0 onset in production (Hombert 1976), and there are crosslinguistic differences in the direction of correlation between F0 and VOT across stop categories within languages (Kingston & Diehl 1994). Speakers of French are known to manipulate vowel F0 in accommodation (Delvaux & Soquet 2007). Among a population of English speakers, an increase in VOT may be expected to be accompanied by an increase in onset vowel F0; however, as this relationship may not be an automatic one, this expectation is a weak one.

Duration of following vowel: A longer VOT is correlated with a longer following vowel duration, due to general pressures for a fixed ratio of segment durations across speech rates (Boucher 2002). Vowel duration is known to be accommodated in other contexts than VOT lengthening (Giles et al. 1991).

3.2 Coincident cues and antagonistic accommodation

I will use stop closure duration to illustrate how coincident cues might illustrate antagonistic accommodation. As previously discussed, stop closure duration is inversely correlated with VOT duration (Lehiste 1970, Boucher 2002). Given a model speech signal with lengthened VOT and average stop closure duration, different predictions are possible regarding the behavior of accommodating speakers as regards stop closure duration. (Assume for now that speakers have the same average closure duration as the model.) We might predict that speakers will shorten their stop closure duration in order to signal the VOT increase being targeted in accommodation to the VOT-adjusted speech signal. Alternatively, we may predict that speakers will increase their VOT but leave stop closure constant. It is not important at this point which of the two features is more likely to be adjusted; the predictions made in either case are formally equivalent in a discussion of the types of accommodation that are possible.

But what if speakers have a shorter average stop closure duration than the model signal, as well as a shorter VOT? In this case, convergence in both features would run counter to the tendency for their inverse correlation. In this instance we have a different set of predictions, shown in Table 1.

Table 1: Possible speaker adjustments for coincident cues across conditions (model with longer VOT and closure than speaker)

	VOT	Closure	Explanation	Result
1.	same	longer	Accommodation to closure, not to VOT	Convergence
2.	longer	same	Accommodation to VOT, not to closure	Convergence
3.	longer	longer	Convergence in both VOT and closure	Convergence
4.	longer	**shorter**	Convergence in VOT, coincident adjustment to closure	Antagonistic
5.	**shorter**	longer	Convergence in closure, coincident adjustment to VOT	Antagonistic

In strategies 1-3, all changes to VOT and stop closure are in the direction of phonetic convergence. However, the resulting adjustments are contrary to the general tendency observed for individual speakers as regards the relationship between the two features. In strategies 4-5, the speaker is exhibiting divergence of one of the two features in question, and convergence of the other. If one of the latter two predictions holds, the result is antagonistic accommodation, which can be attributed to restrictions in the speaker's timing relations.

In the event that speakers have a longer average stop closure duration than the model signal, as well as a shorter VOT, strategies 3-5 are functionally equivalent as far as predicted behavior. If speakers lengthen their VOT in convergence, coincident adjustment to closure duration will be in the same direction as direct accommodation would predict; if they shorten their closures, coincident adjustment to VOT would converge in length.

3.3 Predictions

As the intent of the reported study was to investigate the existence of antagonistic accommodation, a model speaker was used with long closure duration as well as lengthened VOT, in order to make strategies 4-5 available to speakers. Given that previous studies have consistently found convergence in VOT, and the trade-off between VOT and stop closure is well established, it was predicted for this study that speakers would generally pursue strategy 4: they would diverge from the model in closure duration and converge in VOT.

Predictions for speakers' adjustment of vowel duration and onset F0 are contingent on their individual values for these cues. Given that the relationship between F0 and VOT is not clearly an automatic one, speakers are less likely to diverge in F0 as a physiologically inevitable byproduct of convergence in VOT. My prediction then is that they will either not adjust onset F0, or converge in vowel F0, but do so independently of their convergence in VOT. For vowel duration, coincident cue adjustment is predicted. However, some speakers will likely have longer, and others shorter, baseline vowels than the model, meaning that divergence in vowel duration is expected for some but not all speakers.

3.4 Individual differences in coincident cues

As it is possible that individuals will use different strategies in accommodating to their interlocutor during the experiment, it is possible that each of these patterns will be evinced by different speakers. For example, some speakers may adjust onset vowel F0 while others do not. Among those who do, some may converge slightly toward the model talker, while other speakers exhibit hyperconvergence. Should such differences in accommodation be found, they may be attributable to social characteristics of the speakers. Yu (2013) delineates a set of predictions in this vein, proposing a link between personality traits and the degree of compensation for coarticulation evinced in listening tasks. In order to investigate the possibility that a particular speaker's profile of social characteristics might in some way predict the ways in which coincident cues are adjusted in accommodation, the current experiment included a Big Five personality questionnaire (Saucier 1994, following Yu 2013), and a personal empowerment questionnaire (Rogers et al. 1997).[5]

4. Methodology

The experiment consisted of three blocks: (1) baseline recording, (2) repetition of target exposure, and (3) post-exposure recording. Each session typically lasted twenty minutes. Participants were tested individually in a sound booth equipped with a PC, a microphone (AKG C3000), and headphones (AKG K240 Studio). The experimental stimuli were presented using a Python script. Prior to the baseline recording, participants filled out a questionnaire containing biographical data (age, sex, places of residence and languages spoken), a Big Five personality questionnaire, and a personal empowerment questionnaire.

In the baseline recording block, the words in the production list were visually presented on a monitor one at a time. Words persisted on the screen for 3 to 3.1 seconds with the interval in a uniformly random varying distribution, and were then replaced by the subsequent word with no break in between. Participants were given the following instruction: "You will see words. Please say them clearly and quickly." In the target exposure block, the participants were asked to repeat the words that they heard produced by the model talker over headphones. The instruction read: "Now you will hear words. Please repeat them clearly and quickly." No visual stimuli were presented in the exposure block. Finally, in the post-exposure recording block (which was identical to the baseline recording block), the participants were instructed to produce the words in the production list for a second time, providing a post-exposure recording. Across the three blocks, the words were presented in random order for each subject. Participants' tokens were digitally recorded into a computer at a sampling rate of 22,050 Hz.

The production list consisted of 120 words: 104 test words and 16 filler words (see Appendix A). For each word, the target segment was a voiceless stop in the onset of the word's stressed syllable. Among the test words, 49 had initial target voiceless stops (e.g., TENSION), 37 had initial schwa (e.g., ATTENTION), and 18 had initial unstressed syllables with onsets (e.g., DETENTION). Five of the words with initial targets and five of the words with initial schwa had glides following the stop (/k/ in all cases). The same word list was used in both the baseline and post-exposure conditions.

In order to forestall fatigue in participants, the listening list was shorter than the production list, consisting of only 80 words: 65 test words and 15 filler words. Each word was repeated, resulting in a total of 160 list items. Among the test words, 38 had initial

[5]It is also possible that differences in accommodation may be due to individual differences at the level of auditory perception; this study has nothing useful to contribute to such a line of inquiry.

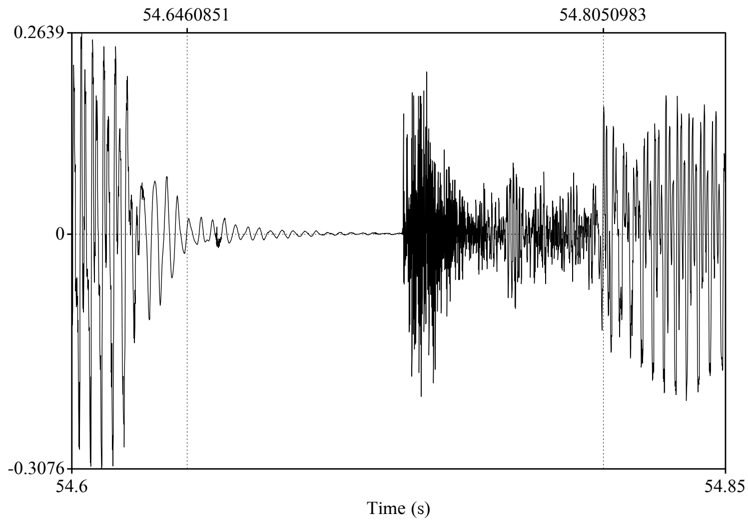
stops, 26 had initial schwa, and 1 had an initial unstressed syllable with an onset (MACAW). One of the words with an initial stop and one of the words with an initial schwa had glides following the stop (CHOIR and ACQUIT). One of the words in the listening list was not on the production list (PASTRY).

A female American English speaker native to the San Francisco Bay Area (in which the experiment was conducted) served as the model talker, and was recorded saying each of the words in the listening list three times. The tokens were digitized at 44,100 Hz. The VOTs for the voiceless stops in the onsets of the stressed syllables for all three tokens of each word were artificially doubled using Praat, such that all parts of the consonant burst and aspiration were extended equally. Only the target segment was manipulated for each token, even if there were other onset voiceless stops in the word. The subjectively most natural token of each word post manipulation was used as a stimulus for the experiment. Overall mean VOT of stressed voiceless consonants in modeled speech was 154.08 ms, with standard deviation 28.52 ms. Mean closure duration for word-medial stops was 61.48 ms. Despite the model's unusually long VOT, in exit interviews participants described the model speaker as sounding natural, albeit hesitant.

Participants were compensated with $5 and course credit. Of the thirty participants, fourteen were female, monolingual English speakers who gave permission to have their recordings used for analysis and discussion. (Only three male participants met all screening criteria; their data was excluded due to this small population size and expected differences between sexes regarding onset vowel F0, one of the cues under investigation.) Subsequently, VOTs and closure durations from the baseline and post-exposure blocks for these fourteen participants were measured from both waveforms and spectrograms using Praat (Boersma & Weenink 2014). Tokens that were pronounced with unexpected stress were excluded from subsequent analysis (n = 6 out of 2,912 target tokens). Due to the difficulties inherent in gauging the onset of stop closure in postpausal position, stop closure durations were only measured for word-medial stops. VOT was measured as the time between the onset of the release burst and the onset of periodic energy due to glottal pulse. Because periodicity was sometimes unclear, the trough before the first unequivocal period was taken as the onset.[6] Closure duration was measured as the time between the initial drop in intensity after periodicity had ceased in the preceding vowel and the onset of the release burst. Figure 3 shows an example of the measurement of overall stop duration (for the word ATTESTED).

[6]This determination has clear ramifications for the calculation of onset vowel F0.

Figure 3: Example of stop duration measurement for the word ATTESTED

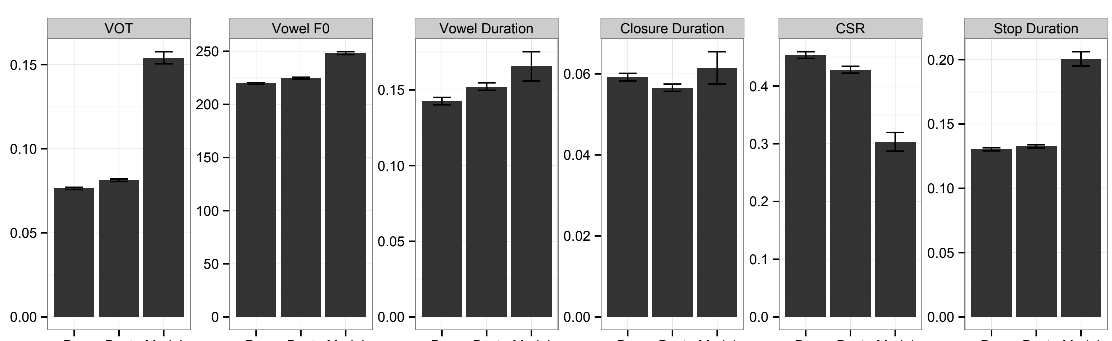


5. Results

5.1 Global results

Table 2 shows a summary of the mean measurements across speakers in the pre-exposure and post-exposure conditions. Across speakers, average VOT increased from the pre-exposure condition to the post-exposure condition. Both vowel F0 and vowel duration also increased across speakers, converging toward the model talker's average for those features. Mean stop closure duration across speakers diverged from the model talker, decreasing despite the model talker's longer mean closure duration. However, the mean CSR (closure-stop ratio) across speakers converged toward the model talker's CSR. Additionally, the mean overall stop duration across speakers did not change significantly (p = 0.158). These data are illustrated in Figure 4.

Figure 4: Global changes across conditions



Vowel F0 is in Hz, Closure-Stop Ratio is as a percentage, all other cues are in seconds.

Table 2: Global changes across conditions

	Condition		Model		ANOVA		Error *p*-values	
	Pre	Post	average	M	*F*	*p*-value	Speaker	Word
VOT (ms)	74.62	79.21	154.08	4.589	68.76	< 0.0001	0.107	0.54
Vowel F0 (Hz)	219.61	225.02	248.18	5.503	72.05	< 0.0001	0.517	0.997
Vowel duration (ms)	137.01	146.83	165.57	9.877	68.06	< 0.0001	0.454	0.838
VOT (ms)†	69.75	74.51	139.15	4.758	39.06	< 0.0001	0.0698	0.611
Closure duration (ms)†	56.74	**53.35**	61.48	3.32	18.48	< 0.0001	0.583	0.936
Closure-stop ratio (%)†	44.60	41.67	30.34	2.82	33.23	< 0.0001	0.123	0.58
Overall stop duration (ms)†	126.58	128.10	200.63	1.43	1.994	0.158	0.269	0.735

All ANOVAs included Condition as a predictor, and Speaker and Word as error terms.
***p*-values refer to significance of change from pre-exposure to post-exposure condition.**
†Does not include word-initial stops.

Taken together, the divergence in absolute closure duration and lack of change in overall stop duration strongly suggest that VOT and closure duration are in an antagonistic relationship as coincident cues. Global convergence in CSR further indicates that divergence in closure duration is a side effect of convergence along other dimensions.

5.2 Results by speaker

Table 3: Individual changes across conditions

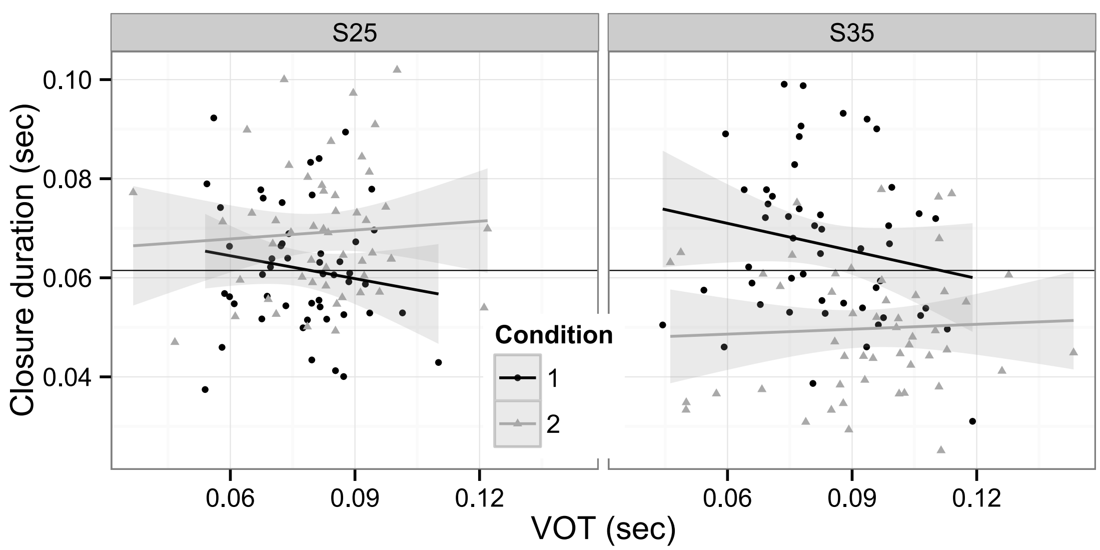
	VOT (ms)		Vowel F0 (Hz)		Vowel duration (ms)		Closure (ms)		CSR (%)		Stop duration (ms)	
Speaker	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Model	154.08 ms		248.18 Hz		165.57 ms		61.48 ms		30.34%		200.63 ms	
S15	76.39	91.50 ***	233.01	236.49	170.38	**184.07**	56.55	**47.78** *	43.88	35.29 ***	129.02	133.23
S17	67.02	69.89	189.69	**189.28**	107.32	113.12 *	52.21	**50.62**	46.44	44.40	112.83	114.73
S20	63.11	**62.55**	230.14	236.84 ***	152.04	155.72	49.72	55.20	43.87	**45.87**	112.15	118.11
S21	83.56	85.82	221.65	253.53 ***	144.73	151.31	55.05	**43.52** **	39.63	32.82 **	134.46	**132.41**
S23	73.81	77.02	219.40	229.44 ***	158.02	160.85	69.65	63.89	50.20	48.72	135.96	**130.50**
S24	71.24	73.13	221.66	225.00	139.47	**138.05**	54.37	**45.86** *	43.74	40.20	122.19	**114.00**
S25	79.05	84.44 ***	201.70	202.40	103.37	111.14 **	61.81	**69.18** ***	44.42	**45.71**	138.94	151.65 ***
S28	70.57	**69.43**	216.05	228.94 ***	108.16	137.32 ***	51.13	62.29 ***	43.19	**48.60** **	119.80	130.17 *
S30	80.95	82.56	229.71	235.12 ***	156.11	156.50	44.49	**36.48** **	35.55	32.63	123.16	**113.20** **
S32	74.33	79.70 *	227.27	231.80 **	118.92	131.63 ***	60.23	**53.63**	47.78	43.61	125.35	**122.78**
S35	92.35	100.41 ***	233.64	237.86	145.91	161.54 ***	66.67	49.73 ***	44.43	34.94 ***	149.94	**143.29**
S39	77.13	89.23 ***	205.47	210.11	134.44	155.10 ***	69.83	58.30 ***	50.00	40.75 ***	139.60	141.99
S40	80.47	83.63	233.94	235.00	158.95	178.46 ***	43.51	48.85	36.39	**38.03**	119.76	128.26 *
S42	54.42	60.23 **	211.09	**198.29** ***	120.19	120.64	58.76	61.37	54.20	51.94	109.65	118.34

Features exhibiting divergence are bold and shaded. Features exhibiting hyperconvergence are in dashed boxes. Asterisks indicate significance levels of t-tests across conditions after Bonferroni correction for tests within speakers. *$p < 0.05$, **$p < 0.01$, ***$p < 0.005$.

Table 3 shows a summary of the mean measurements for each condition broken out by speaker. For each feature under investigation, up to half of the participants diverged from the model talker. All participants had shorter mean VOT and lower mean vowel onset F0 in the pre-exposure condition than the model talker. One participant had a longer mean vowel duration than the model talker (S15) prior to exposure; four participants had longer mean closure durations than the model talker prior to exposure (S23, S25, S35, S39).

All but three speakers (S17, S24, S42) converged significantly in at least one feature. Two speakers (S25 and S42) diverged significantly in a feature – closure duration and onset vowel F0, respectively. Three speakers converged significantly in closure duration to the point of hyperconvergence; of these three, one had a shorter mean baseline closure duration than the model talker (S28), and two had longer mean durations (S35, S39). One example each of antagonistic accommodation and hyperconvergence in stop closure duration are illustrated in Figure 5. Only one speaker (S25) showed significant adjustment of overall stop duration across conditions.

Figure 5: Antagonistic accommodation (S25) and hyperconvergence (S35) in closure duration



S25 diverged in closure duration from the model (*p* < 0.0001) and S35 hyperconverged to the model (*p* < 0.0001). Lines show best-fit linear models for stop closure duration in baseline (Condition 1) and post-exposure (Condition 2) word-medial stops.

Table 3 shows individual changes in VOT for consonants in all positions in the word, as data from stop-initial consonants were incorporated into analysis on vowel duration and vowel F0. Given that the consonantal cues (closure duration, CSR, and overall stop duration) were only measured for word-medial consonants, Table 4 shows VOT changes just for consonants in these contexts.

Table 4: Individual changes across conditions (word-medial stops)

	VOT (ms)		Closure (ms)		CSR (%)		Stop duration (ms)	
Speaker	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Model	139.15 ms		61.48 ms		30.34%		200.63 ms	
S15	72.47	85.45 ***	56.55	**47.78** *	43.88	35.29 ***	129.02	133.23
S17	60.62	64.10	52.21	**50.62**	46.44	44.40	112.83	114.73
S20	62.43	62.91	49.72	55.20	43.87	**45.87**	112.15	118.11
S21	79.41	88.89 *	55.05	**43.52** **	39.63	32.82 **	134.46	**132.41**
S23	66.31	66.60	69.65	63.89	50.20	48.72	135.96	**130.50**
S24	67.83	68.15	54.37	**45.86** *	43.74	40.20	122.19	**114.00**
S25	77.13	82.47 *	61.81	**69.18** ***	44.42	**45.71**	138.94	151.65 ***
S28	68.67	67.88	51.13	62.29 ***	43.19	**48.60** **	119.80	130.17 *
S30	78.67	**76.73**	44.49	**36.48** **	35.55	32.63	123.16	**113.20** **
S32	65.12	69.15	60.23	**53.63**	47.78	43.61	125.35	**122.78**
S35	83.27	93.55 ***	66.67	49.73 ***	44.43	34.94 ***	149.94	**143.29**
S39	69.78	83.69 ***	69.83	58.30 ***	50.00	40.75 ***	139.60	141.99
S40	76.25	79.42	43.51	48.85	36.39	**38.03**	119.76	128.26 *
S42	50.89	56.97	58.76	61.37	54.20	51.94	109.65	118.34

Features exhibiting divergence are bold and shaded. Features exhibiting hyperconvergence are in dashed boxes. Asterisks indicate significance levels of t-tests across conditions after Bonferroni correction. *p < 0.05, **p < 0.01, ***p < 0.005.

After recording the stimuli, the model talker indicated that she had guessed the experiment had to do with the presence or absence of initial schwa in the various words. Intons-Peterson (1983) indicates that the effect of presented stimuli can be affected by the expectations of the experimenter, or in this case interlocutor. The model talker's reported focus on the context before the target consonants may be a contributing factor to the significance of Stop Position or Multisyllabicity as a predictor for each of the cues investigated.

5.3 Results by cue

5.3.1 VOT

As expected, most participants had a longer mean VOT in the post-exposure condition, with six participants showing a statistically significant increase ($p < 0.05$). In absolute terms, two participants diverged from the model VOT across all words (S20 and S28), while a third diverged specifically in the word-medial context (S30). This differing behavior depending on the environment of the stop, while not approaching the level of significance, indicates that a blanket assertion of convergence in VOT may be an oversimplification.

A mixed effects regression model predicting VOT was fit with Block (either *pre-exposure* or *post-exposure*), Vowel Identity, Stop Identity, Stop Position within the word, and following Vowel Onset F0 as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block made significant contributions to model likelihood (see Table 5). The resulting model had an adjusted R^2 of 0.5523. Within-block Trial Position and following Vowel Duration did not prove to be significant predictors, nor did interactions between Stop Identity and Stop Position or between Block and Trial Position. Vowel F0 being a significant predictor is likely due to a bias in the placement of the boundary between consonant and vowel; dropping Vowel F0 as a predictor gave the model an adjusted R^2 of 0.5509.

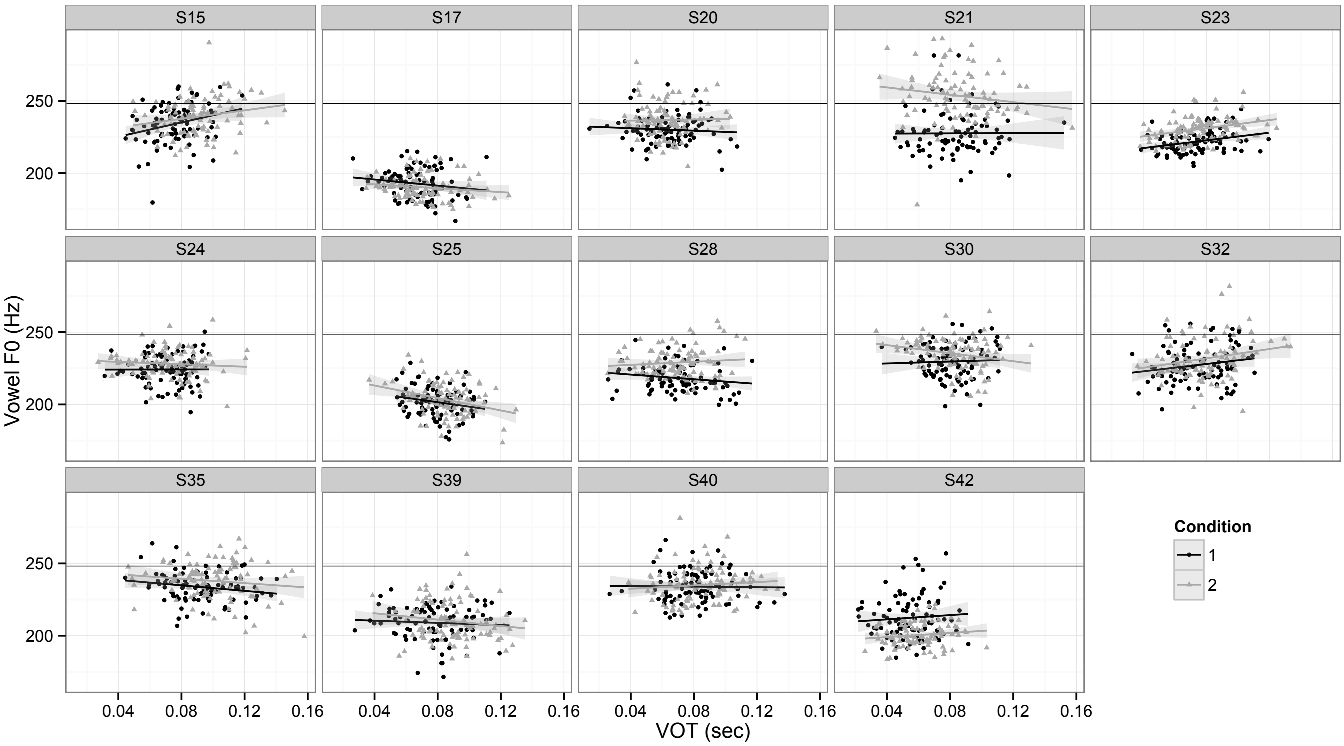
Table 5: Random slopes by speaker for VOT across conditions

	Slope (ms)		Slope (ms)
S15	***12.511	S28	0.256
S17	2.778	S30	3.172
S20	0.429	S32	5.241
S21	5.105	S35	*8.492
S23	4.399	S39	***10.048
S24	2.283	S40	3.670
S25	5.501	S42	3.620

5.3.2 Vowel onset F0

Overall, twelve of the fourteen participants converged toward the model's average onset F0, which was higher than that of all participants in the baseline condition. Two participants (S17 and S42) diverged in absolute average F0, including the single participant with the lowest baseline F0 (S17). One participant hyperconverged in F0 (S21). Best-fit lines for F0 as a function of VOT are shown in Figure 6 for each participant by condition.

Figure 6: Vowel F0 by speaker



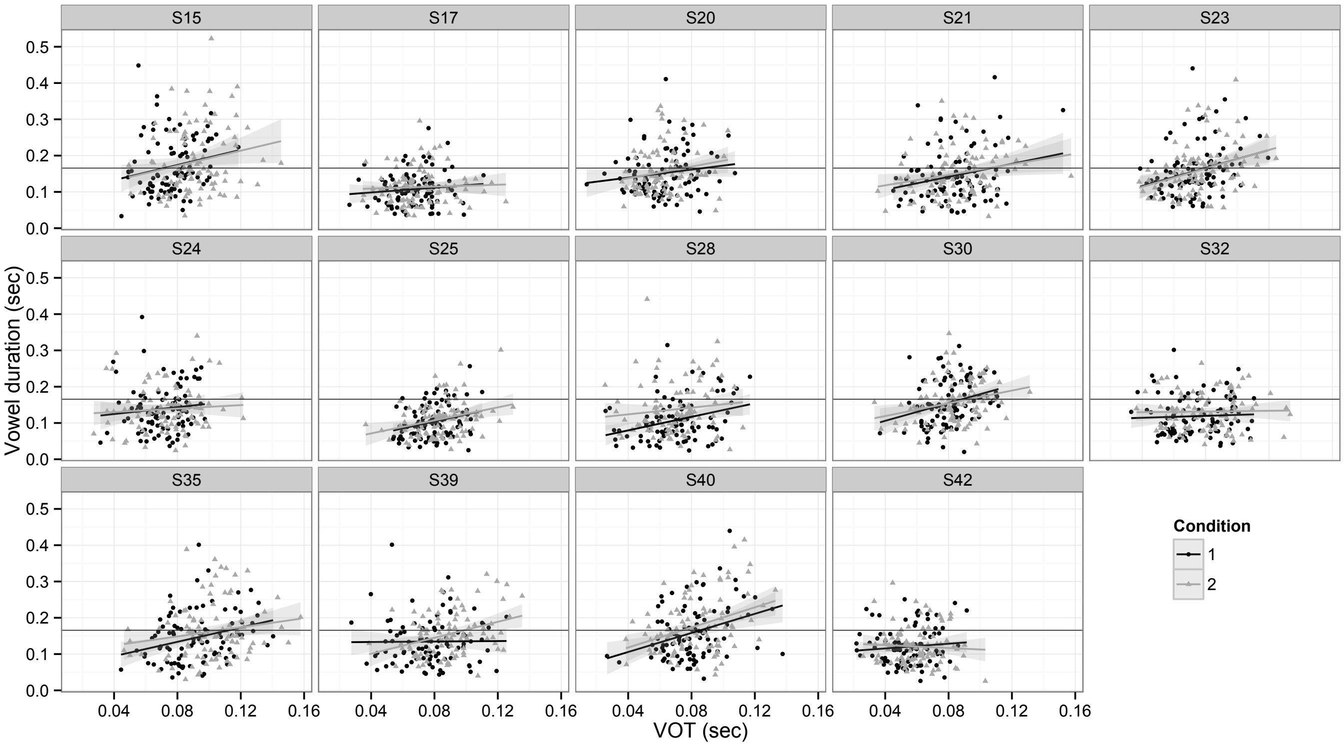
Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average onset vowel F0 of 248.18 Hz is shown for comparison.

A mixed effects regression model predicting vowel onset F0 was fit with Block, Vowel Identity, Stop Position within the word, and VOT duration as predictor variables. VOT being a significant predictor indicates that adjustment of F0 was automatic, contrary to prediction. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood (R^2 = 0.5572). Stop Identity and Vowel Duration were not significant predictors, nor was Trial Position on its own. The interaction between Stop Position and Stop Identity also did not prove significant. Multisyllabicity (whether the word is monosyllabic or multisyllabic) was a worse predictor than Stop Position as a whole, and so was not included.

5.3.3 Vowel duration

Overall, twelve of the fourteen participants converged toward the model's average onset vowel duration. Two participants (S15 and S24) diverged in absolute average vowel duration. One participant hyperconverged in vowel duration (S40). Best-fit lines for vowel duration as a function of VOT are shown for each participant by condition in Figure 7.

Figure 7: Vowel duration by speaker



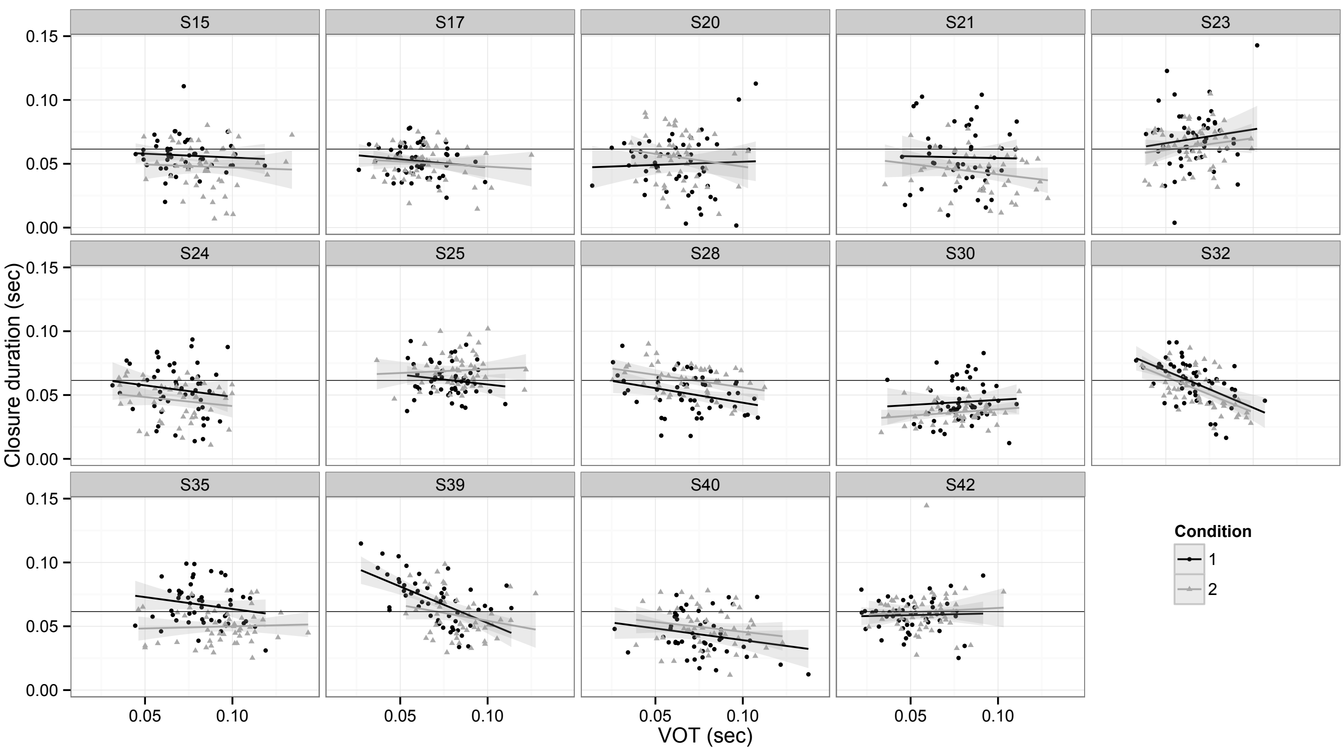
Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average vowel duration of 165.57 ms is shown for comparison.

A mixed effects regression model predicting vowel duration was fit with Block, Vowel Identity, Following Segment Type (voiced stop, voiceless stop, voiced fricative, liquid, nasal, vowel, or no following segment), Multisyllabicity, and within-block Trial Position as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood, although Trial Position lost significance after Bonferroni correction (R^2 = 0.8197 with by-speaker Block). Stop position within the word and the identity of the following consonant did not prove to be significant predictors. Despite its strong correlation with vowel duration, VOT also was not a significant predictor for vowel duration, either by itself or in interaction with task condition. Since all but one speaker had pre-exposure vowels that were shorter than the model talker's vowels on average, this may indicate that speakers were accommodating to vowel duration independently from accommodation to VOT. While Multisyllabicity was a significant predictor of vowel duration, an interaction between Multisyllabicity and Block was not significant, indicating that changes to vowel duration were independent of the model talker's perceived speech rate.

5.3.4 Stop closure duration

Best-fit lines for closure duration as a function of VOT are shown for each participant by condition in Figure 8. Speakers with both longer (S25) and shorter (S15, S21, S24, S30) average closure durations than the model diverged in closure duration. This result indicates that divergence is not due to an intrinsic bias toward longer or shorter closure durations in themselves; moreover, it strongly suggests that this divergence is a side effect of convergence in other features.

Figure 8: Stop closure duration by speaker



Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average stop closure duration of 61.48 ms is shown for comparison.

A mixed effects regression model predicting stop closure duration was fit with Block, Trial Position, Stop Position within the word, Stop Identity, and VOT duration as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood (R^2 = 0.4771). Across speakers, the model found global divergence in closure duration from the model talker of 2.64 milliseconds on average (β = –2.641 ms, standard error = 0.7768, t = –3.399).

6. Discussion

In accommodating to a model talker with artificially lengthened VOT, speakers were expected in the aggregate to converge toward model VOT, and in so doing diverge from the model talker in stop closure duration. Predictions for the other targeted coincident cues of VOT were contingent on individual speakers' values for those cues. Coincident cue adjustment was entertained but not expected for onset vowel F0, as the relationship between that cue and VOT was hypothesized to be a weak one. Following vowel duration was expected to undergo coincident cue adjustment, although not all speakers would have vowels that were shorter or longer than the model.

6.1 General findings

Of the three coincident cues to VOT under investigation – stop closure duration, onset vowel F0, and vowel duration – only the first two are significantly predicted by VOT across conditions. All three, however, change significantly after exposure to an external speech signal. This indicates that speakers accommodated independently to both VOT and vowel duration. Best models for vowel F0 and closure duration include Trial Position effects by speaker. This behavior is predicted by exemplar theories of phonetic organization (Johnson 1997, Pierrehumbert 2001). Best-fit models for vowel duration and closure duration also include Trial Position effects as fixed effects; while it is possible that these effects are those predicted by exemplar theories, the fact that these models were not improved with by-speaker effects suggests that they may be instead attributable to fatigue or neutralization over the course of the experiment.

Vowel F0 was significantly predicted by VOT across conditions. This coincident adjustment runs contrary to the hypothesis that vowel F0 would be controlled and adjusted independently of VOT. Also contrary to expectation, vowel duration was not significantly predicted by VOT across speakers, although half of speakers exhibited significant convergence to model vowel duration.

Closure duration is an interesting case. Exactly half of speakers exhibited absolute convergence in average closure duration, although only three did so with any statistical significance. Convergence occurred for speakers with mean pre-exposure closure durations that were both shorter than the model talker (S20, S28, S40, and S42) and longer than the model talker (S23, S35, S39). Likewise, divergence occurred for speakers with both shorter (S15, S17, S21, S24, S30, S31) and longer (S25) pre-exposure closure durations than the model talker. The fact that speakers diverged from the model closure duration both by lengthening and shortening their own closures indicates that this divergence is an instance of antagonistic accommodation.

6.2 Individual differences

Table 6 is a correlation matrix of the changes to measured acoustic features measured across conditions and the results of the personality and power questionnaires (Rogers et al. 1997; Saucier 1994) administered before the experiment. No correlation reached $p < 0.10$ after Bonferroni correction. Given the relatively small number of speakers analyzed, this lack of significance is not surprising.

Table 6: Correlations of characteristics and changes across conditions

	ΔVOT	ΔF0	ΔVowelDur	ΔClosureDur	ΔCSR	ΔStopDur
Power	0.321	-0.348	-0.577	0.155	0.138	-0.111
Extraversion	0.190	-0.034	-0.365	0.398	0.206	0.055
Agreeableness	0.239	-0.040	-0.109	0.456	0.252	-0.003
Conscientiousness	0.198	-0.563	-0.088	-0.055	0.136	-0.137
Neuroticism	0.228	-0.414	-0.573	-0.034	0.093	0.247
Openness	0.082	-0.399	-0.068	0.256	-0.038	0.137

However, it can be asserted that different participants behaved differently. Of the six speakers with a significant change to VOT and/or closure duration across conditions, two converged in both VOT and closure duration (S35, S39), two converged in VOT without adjusting closure duration (S15, S21), and two adjusted closure duration without adjusting VOT (S25, S28). While S25 had an absolute average closure duration that was longer than that of the model talker, a two-tailed t-test did not show a significant difference in closure duration ($t = -1.0054$, df = 73.357, $p = 0.318$). Empirically, both S25 and S28 lengthened their closure duration; it cannot be stated with assurance whether they both converged in this regard.

6.3 Automaticity of coincident cues

Varying accounts of phonetic organization make differing predictions regarding how automatic the adjustment of coincident cues is. Kingston & Diehl's (1994) *phonetic reorganization* account holds that speakers are able to control cues of phonemic categories independently, as demonstrated by their ability to vary the phonetic realization of speech sounds between contexts. Phonetic implementations are "capacity-limited, attention-demanding, relatively easily learned and modified, and often accessible to conscious inspection" (Kingston & Diehl 1994). The phonetic reorganization account states that while some coincident cues are automatic due to physiological constraints on speakers' articulation, most coincident cues are controllable, including many of those that are phonologized in a language. This account predicts that controllable cues will not be automatically adjusted in concert with VOT; they will instead be independently adjusted in relation to the received speech signal's explicit value for those cues. In the context of this study, the phonetic reorganization account predicts that vowel F0 and vowel duration will not be automatically adjusted, whereas closure duration will. Although vowel duration was indeed adjusted independently of VOT, vowel F0 was not. In the event that there is a physiological link between F0 and VOT, these findings are entirely consistent with a phonetic reorganization account.

A contrary set of predictions is made by an *exemplar theory* account of phonetic organization (Johnson 1997; Pierrehumbert 2001). Exemplar theory holds that different subsets of experience are called on in deciding which variant of a given linguistic structure to use. Under this account, all coincident cues to VOT in English should be adjusted in accommodation to a signal with lengthened VOT, since all coincident cues are by definition associated in the main with VOT lengthening. This adjustment should take place even if the model for accommodation falls outside of the set of all prior experiences for the speaker, assuming that the speech signal is still analyzed as belonging to the same set of experiences. The most robust cues of VOT should evince the greatest degree of accommodation, due to speakers' increased familiarity with those cues (everybody uses them). These predictions appear to be borne out to some extent in the results of this study, as closure duration and vowel F0 were both automatically adjusted along with VOT. The lack of coincident adjustment of vowel duration may be attributed to the relatively large variation evinced by the model (sd = 76.5 ms).

6.4 Consonantal vs. vocalic accommodation

As has been alluded to, onset vowel F0 and vowel duration might properly be considered cues to the vowel following the manipulated consonant, rather than direct cues to the consonant itself. Only half of the speakers evinced significant levels of accommodation to both consonantal and vocalic cues. Most previous accommodation studies have restricted their area of inquiry one or the other category: consonantal cues are generally easier to quantize, whereas vocalic cues are perhaps easier to perceive. The results of this experiment indicate that the conflation of these two classes of targets for accommodation may lead to erroneous conclusions about how much accommodation takes place.

6.5 How to measure accommodation

In any study investigating the effects of sociological factors such as interlocutor attractiveness or likeability, it would not be enough to measure phonetic convergence simply in terms of VOT. As these findings indicate, participants S20 and S28 would not be seen as converging in such a study. However, it is not at all clear that they diverged, or even that they failed to show convergence: both speakers converged in vowel duration, and S28 also converged in vowel F0 and in closure duration. Given that speakers generally seem not to be able to independently manipulate VOT and closure duration at the same time, it is misleading to expect convergence along every feature – and in a study examining only one feature, that feature may well be the "wrong one" for some participants. Of course, this understanding leads to the question: is total divergence even possible? If enough features are measured, participants will inevitably exhibit convergence in at least one. For a speaker in this study to exhibit total divergence, they would have to either shorten their mean VOT and lengthen an already-long closure duration, or have an uncommonly long VOT pre-exposure. Nielsen (2011) found that speakers would not shorten VOT in voiceless stops in accommodation to a model signal with shortened VOT. She suggested that this might be due to the possibility of introducing phonological ambiguity, given that voiced and voiceless stops in English are differentiated primarily by the comparatively longer VOT in voiceless stops. This finding indicates that there are circumstances in which total divergence will not happen. But this does not mean that quantitative measurement of accommodation is straightforward.

On the other hand, qualitative measurement of accommodation is no more straightforward. AXB similarity tasks used to qualitatively confirm accommodation (cf. Goldinger 1998, Nye & Fowler 2003) will not disambiguate between the adjustment of different cues to accommodation. It is likely that some cues are more salient in some fashion to speakers than others. However, the comparative salience of a feature may not be universal across a population. Given the robustness of VOT's accommodation effects and its crucial role in English phonology, it is sensible to expect VOT to be one of these salient cues. In that light is surprising that as many speakers displayed significant adjustment to closure duration as did to VOT. It may be possible to look at whether speakers who do not accommodate to VOT evince this lack of salience in their production of stops – for example, they may invariably voice phonologically voiced stops whereas other American English speakers do not.

6.6 Ramifications for sound change

If – as is often assumed – convergence is the default state of affairs in interlocution, antagonistic accommodation may be an inevitable byproduct thereof. As such, it is conceivable to delineate a course for sound change in which no misperception is necessary on the part of the listener. Antagonistic accommodation is analogous to Ohala's formulation of hypercorrection (1989, 1993), in that it is an "inappropriate application of [...] corrective rules" (Ohala 1989): speakers making attendant adjustments to their production of a received target. The key difference is that the 'correcting' is not due to a mismatch between a speaker's intended production and a listener's perception of that intention, rather between a speaker's intended production and their ability to effect that production.

References:

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. (2011). Effects of speaker evaluation on phonetic convergence. In *Proceedings of the International Congress of the Phonetic Sciences*, Hong Kong, 192-195.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society* 39(4), 437-456.

Bock, J.K. (1986). Syntactic persistence in language production. *Cognitive Psychology* 18, 355-387.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer. Version 5.3.62, retrieved 2 January 2014 from http://www.praat.org/.

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics* 64 (1), 121-130.

Bourhis, R.Y. & Giles, H. (1977). The language of intergroup distinctiveness. In *Language, ethnicity, and intergroup relations*. Howard Giles, ed. New York: Academic Press.

Chang, C.B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40(2), 249-268.

Costa, P.T., Jr. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources.

Delvaux, V. & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145-173.

Fowler, C.A., Brown, J., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* 49, 396-413.

Gentilucci, M., & Bernardis, R. (2007). Imitation during phoneme production. *Neuropsychologia* 45, 608-615.

Giles, H. (1971). *A study of speech patterns in social interaction: Accent evaluation and accent change*. Unpublished Ph.D thesis, University of Bristol.

Giles, H., Coupland, J., and Coupland, N. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of Echoes? An episodic theory of lexical access. *Psychological Review* 105(2), 251-279.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11(4), 716-722.

Gregory, S.W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior* 14, 237-251.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?. *The Journal of the Acoustical Society of America* 109(2), 764-774.
Hombert, J. M. (1976). The effect of aspiration on the fundamental frequency of the following vowel. *Proceedings of the 2nd Annual Meeting of the Berkeley Linguistics Society*, 212-219.
House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1), 105-113.
Intons-Peterson, M.J. (1983). Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception and Performance* 9(3): 394-412.
Jaffe, J., & Feldstein, S. 1970. *Rhythms of dialogue*. New York: Academic Press.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, 145-165.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language* 70(3), 419-454.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge: Cambridge University Press.
Lisker, L. (1986). "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1), 3-11.
Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from a shadowing task. *Cognition* 109, 68-173.
Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32, 790-804.
Nielsen, K. (2008). *The specificity of allophonic variability and its implications for accounts of speech perception*. Los Angeles: University of California, Los Angeles dissertation.
––. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2), 132-142.
Nye, P. W. & Fowler, C. A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics* 31(1), 63-79.
Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, 173-198.
––. (1993). The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and Perspectives*. 237-278.
Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4), 2382-2393.
Pardo, J. S., Jay, I. C., & Krauss, R.M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics* 72(8), 2254-2264.
Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*, 137-157. Amsterdam: John Benjamins.
Rogers, E. S., Chamberlin, J., Ellison, M. L., & Crean, T. (1997). A consumer-constructed scale to measure empowerment among users of mental health services. *Psychiatric services* 48(8), 1042-1047.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Co.

Shockley, K., Sabadini, L., and Fowler, C.A. (2004). Imitation in shadowing words. *Perception & Psychophysics* 66(3), 422-429.

Weinreich, U., Labov, W., & Herzog, M.I. (1968). *Empirical foundations for a theory of language change*. University of Texas Press.

Yu, A.C.L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: implications for models of sound change. *PloS one* 5(8), e11950.

––. (2013). Individual differences in socio-cognitive processing and sound change. *Origins of Sound Patterns: Approaches to phonologization*. Oxford University Press.

Appendix A: Table of stimulus word frequency.

Word	Freq	LgFreq	Word	Freq	LgFreq	Word	Freq	LgFreq
a cappella	0.43	1.3617	cord	7.02	2.5551	pend*		
accompany	4.75	2.3856	cost	54.92	3.4475	petunia	2.08	2.0294
accord	1.63	1.9243	count	89.96	3.6617	picante	0.04	0.4771
accost	0.06	0.6021	cues	0.69	1.5563	picard	1.53	1.8976
account	44.71	3.3581	cult	4.45	2.3579	pinion	0.16	0.9542
accuse	5.69	2.4639	curd	0.43	1.3617	point	236.53	4.0815
accustom	0.12	0.8451	currents	1.69	1.9395	posable	0.02	0.301
acquaint	0.39	1.3222	custom	6.20	2.5011	potato	11.29	2.7612
acquire	2.65	2.1335	cute	87.75	3.6509	potential	18.82	2.9827
acquit	0.47	1.3979	department	63.84	3.5128	quaint	2.18	2.0492
acute	2.94	2.179	detention	6.53	2.5237	quit	90.10	3.6624
akin	0.27	1.1761	earache	0.29	1.2041	recognize	34.31	3.2433
all	5161.86	5.4204	else	449.16	4.36	repent	2.41	2.0934
apace	0.06	0.6021	interest	50.94	3.4148	retaining	0.65	1.5315
apart	47.02	3.38	katana	0.08	0.699	risk	49.04	3.3983
apiece	3.96	2.3075	kin	4.27	2.3404	satirical	0.18	1
appal	0.04	0.4771	kosher	2.69	2.1399	standard	18.43	2.9736
apparent	4.22	2.3345	look	1947.27	4.997	tack	2.12	2.0374
appeal	13.00	2.8222	macaw	0.24	1.1139	tall	32.33	3.2175
appearing	2.33	2.0792	marriage	77.06	3.5945	target	37.96	3.2871
appease	0.49	1.415	memory	48.57	3.3941	taskmaster	0.16	0.9542
append*			natural	42.35	3.3347	tasty	6.31	2.5092
appoint	2.04	2.0212	nose	69.75	3.5512	tempt	2.53	2.1139
Atari*			occult	1.57	1.9085	tend	12.27	2.7973
atone	0.84	1.6435	occurred	14.45	2.8681	tension	8.55	2.6405
attack	75.55	3.5859	occurrence	1.18	1.7853	tested	10.53	2.7308
attempt	19.12	2.9894	opinion	42.00	3.331	tiger	18.53	2.9759
attend	14.02	2.8549	opposable	0.35	1.2788	tire	12.37	2.8007
attention	98.67	3.7018	pace	9.57	2.6893	tizzy	0.18	1
attested	0.12	0.8451	pall	0.45	1.3802	tofu	2.69	2.1399
attire	1.49	1.8865	panther	2.57	2.1206	tomato	5.90	2.48
attorney	40.39	3.3141	paper	103.35	3.722	tone	16.86	2.935
attune	0.02	0.301	parent	13.14	2.8267	torque	0.73	1.5798
battalion	5.84	2.4757	part	261.51	4.1251	tourney	0.14	0.9031
botanical	1.04	1.7324	pastry†	1.92	1.9956	tundra	0.27	1.1761
Capone	1.63	1.9243	pawn	4.33	2.3464	tune	15.61	2.9015
catastrophe	2.47	2.1038	peace	69.61	3.5504	upon	62.73	3.5051
catatonic	0.78	1.6128	peas	4.65	2.3766	welfare	7.88	2.6053
choir	5.31	2.4346	peel	5.35	2.4378	wish	235.12	4.0789
company	147.20	3.8755	peering	0.33	1.2553	yes	1996.76	5.0079
Copernicus	0.67	1.5441						

*Does not occur in the SUBTLEX-US corpus.
†Did not occur in the production list.