



## Accommodating variation: Dialects, idiolects, and speech processing

Tanya Kraljic <sup>a,\*</sup>, Susan E. Brennan <sup>b</sup>, Arthur G. Samuel <sup>b</sup>

<sup>a</sup> *Center for Research in Language, University of California at San Diego, La Jolla, CA 92093-0526, United States*

<sup>b</sup> *Department of Psychology, State University of New York at Stony Brook, Stony Brook, NY 11794-2500, United States*

Received 16 October 2006; revised 25 July 2007; accepted 26 July 2007

---

### Abstract

Listeners are faced with enormous variation in pronunciation, yet they rarely have difficulty understanding speech. Although much research has been devoted to figuring out how listeners deal with variability, virtually none (outside of sociolinguistics) has focused on the source of the variation itself. The current experiments explore whether different kinds of variation lead to different cognitive and behavioral adjustments. Specifically, we compare adjustments to the same acoustic consequence when it is due to context-independent variation (resulting from articulatory properties unique to a speaker) versus context-conditioned variation (resulting from common articulatory properties of speakers who share a dialect). The contrasting results for these two cases show that the source of a particular acoustic–phonetic variation affects how that variation is handled by the perceptual system. We also show that changes in perceptual representations do not necessarily lead to changes in production.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Speech perception; Speech production; Perceptual learning; Coarticulation; Dialects; Idiolects; Language change; Speech accommodation

---

---

\* Corresponding author.

*E-mail address:* [tkraljic@crl.ucsd.edu](mailto:tkraljic@crl.ucsd.edu) (T. Kraljic).

## 1. Introduction

Psycholinguists have known for a half century that pronunciation varies for different reasons. For example, Liberman (1957) showed that the acoustics of stop consonants are drastically influenced by adjacent vowels, and Peterson and Barney (1952) showed that the acoustic realization of any particular vowel is influenced by the age, gender, and dialect of the speaker.

Some sources of variation are due to unusual articulatory properties of a particular speaker, such as lisps or other speech impediments, or temporary physical states such as speaking with something in one's mouth or being drunk. Other types of variation reflect more general articulatory changes that result from adapting a sound to the phonetic context; such contextually-driven changes are commonly found in dialects: Boston speakers drop their /r/'s only when they occur after vowels (e.g., *pahk* for *park*); American English speakers 'flap' /t/s that occur intervocalically at the onset of an unstressed syllable (e.g., the second /t/ in *total*) but they never flap the first /t/; and in the American South, /z/ often becomes [d]-like when it occurs before /n/ (for example, [bɪdnɪs] for *business*). In these and other ways, pronunciation can be highly variable.

The fundamental question in speech perception is this: How do listeners arrive at a stable percept from a continuously variable signal, especially in the face of changes in pronunciation from so many different sources? Every account of speech perception must deal with the question of acoustic variability somehow, whether by regarding it as a source of noise to be filtered out en route to recovering phonetic features (see Tenpenny, 1995, for a review), or by encoding it as an aid to subsequent perception (Goldinger, 1998; Goldinger, Kleider, & Shelley, 1999; Nygaard & Pisoni, 1995; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1993; Remez, Fellowes, & Rubin, 1997). But in this endeavor, a basic and important question has been largely overlooked: Might successful speech perception rely on different processing for different types of variation? Consider the old joke:

Tourist: *Excuse me, is it pronounced 'Hawaii' or 'Havai'?*

Benny Hill: *Havaii*

Tourist: *Thank you!*

Benny Hill: *You're velcome!*

When Benny Hill pronounces /w/ as [v] in Hawaii, we first attribute the variation as lexical; however once we hear him pronouncing /w/ as [v] in a second and unrelated context (*You're welcome*), it becomes obvious that this attribution is wrong: The variation reflects a non-contextual idiosyncrasy of the speaker. But current accounts of speech perception do not distinguish contextually-driven from non-contextually-driven pronunciations; they treat *variation as variation*.<sup>1</sup>

<sup>1</sup> Some accounts have, however, distinguished specific types of variation on the basis of acoustic characteristics; see Gow and Im, 2004.

Even when two variants are acoustically identical, they may have different possible causes, which should have consequences for how they can be processed. In the present experiments, we test the hypothesis that variants with the same form have different processing consequences depending on why they occur. We use the pronunciation  $\sim sf$ , which is ambiguous between [s] and [ʃ]. Recent studies have shown that when a listener hears idiolectal  $\sim sf$ , the exposure leads to a perceptual ‘retuning’ whereby the listener’s phonemic representation of /s/ is expanded to accommodate the new pronunciation (Kraljic & Samuel, 2005; also see Norris, McQueen, & Cutler, 2003 for another idiolectally produced contrast). What is unclear is whether such learning depends on the idiosyncratic, non-contextual nature of the variation in those studies, or whether it occurs more generally, with all variation treated alike.

Fortuitously,  $\sim sf$  affords a natural opportunity to study the same pronunciation as both non-contextual and contextual: Some communities of English speakers pronounce /s/ as nearly [ʃ], but only when the /s/ is followed by [tr] (as in *street*, or *astronaut*). In this case  $\sim sf$  is contextually-driven; in particular, this variant of /s/ is due to coarticulation: /s/ is articulated with a retracted tongue position, anticipating the /r/ in the subsequent [tr] cluster. This dialectal realization of  $\sim sf$  is well established in the Philadelphia English dialect (Labov, 1984), but is not entirely regional; it appears to be part of a sub-dialect<sup>2</sup> produced consistently by some but not all speakers within the regions it occurs (e.g., see Baker, Mielke, & Archangeli, 2006). These regions include Hawaii, New Zealand, Australia, and the American South (e.g., see Lawrence, 2000; Shapiro, 1995), but not the West Coast. Critically for our purposes, this variant is extremely common in speakers of New York/Long Island English. The (acoustically identical) idiolectal variant  $\sim sf$ , in contrast, is not contextually-constrained, but is usually due to some physical state (e.g., having a tongue piercing or to some other idiosyncrasy of the speaker). For ease of exposition, we will refer to the contextually-constrained variant of /s/ as dialectal  $\sim sf$ , and to the non-contextually-constrained variant as idiolectal  $\sim sf$ .

As we noted, dialectal  $\sim sf$  results from coarticulation of [s] with the following [tr]. As such, it could potentially be resolved by assigning features that contribute to the variability to the surrounding phonetic context. For example, when *street* is pronounced as  $\sim sfstreet$ , a listener might uncover the underlying /s/ by assigning the ambiguous features of  $\sim sf$  to the following [tr] (see Gow, 2003; Gow & Im, 2004 for details). Because idiolectal  $\sim sf$  is not contextually-driven, such a strategy would not work; in this case, perceptual learning enables the system to restructure its phonemic representation to accommodate the new pronunciation.

Our hypothesis is that perceptual learning, as the more radical solution, will only occur when the system has no alternative resolution for a variation. When an alternative (such as feature parsing, for example) is available, it would make sense for the system to avoid taking the drastic step of restructuring representations. Thus, the same acoustic variant should lead to different processing consequences.

<sup>2</sup> Although the social correlates of this sub-dialect are not well understood, its use may be more pronounced in “tough” or “colloquial” speech, which suggests a social class factor in some regions.

Exploring variation that occurs dialectally raises another interesting question: How might hearing a particular variation affect subsequent pronunciation? We know that people speak differently depending on their addressees. During a conversation, a speaker's accent and other aspects of utterances often become more similar to those of their interlocutors (e.g., Bell, 1984; Coupland, 1984; Coupland & Giles, 1988; Giles, 1973a, 1973b; Giles & Powesland, 1975; Schober & Brennan, 2003); more long-term, a talker's speech may become more similar to that of the larger community (e.g., Flege, 1987; Gallois & Callan, 1988; Sancier & Fowler, 1997). Our secondary aim is to begin to explore effects of perceptual learning on subsequent production. Does perceptual learning of a pronunciation compel a subsequent change in production of that pronunciation (as, e.g., predicted by Garrod & Anderson, 1987; Pickering & Garrod, 2004)?

## 2. Experiment 1

The purpose of Experiment 1 was to examine whether an identical variant of /s/ will have different processing consequences as a function of its cause. During an initial Exposure phase, we exposed half of the participants to a speaker in which the ~sɹ variant of /s/ was contextually-driven: The pronunciation occurred only when /s/ immediately preceded [tr] (as in *district*, for example) (dialectal ~sɹ). We exposed the other participants to the same ambiguous ~sɹ segment, but it was not context-specific: The pronunciation occurred anywhere an /s/ would normally appear, in words such as *hallucinate* and *pedestal*, and therefore was necessarily attributed to an idiosyncrasy of the speaker (idiolectal ~sɹ).<sup>3</sup> We then measured the /s/-/ʃ/ boundary for each participant to determine whether perceptual learning had occurred. Given our previous work on perceptual learning using idiosyncratically ambiguous fricatives (Kraljic & Samuel, 2005), we expected to find robust perceptual learning for idiolectal ~sɹ. The question was whether we would also see perceptual learning for the group exposed to the dialectal variant. If the perceptual system treats contextually-driven and speaker-driven variation in the same way, we should find perceptual learning for dialectal ~sɹ. However, if the perceptual system represents speaker-driven variation and resolves contextually-driven variation without resorting to representational adjustment, the dialectal variant should not result in perceptual learning.

### 2.1. Method

#### 2.1.1. Participants

One hundred thirteen undergraduate psychology students from the State University of New York at Stony Brook participated for research credit in a psychology

<sup>3</sup> Note that to keep the dialect and idiolect sets as distinct as possible, words with [str] cluster were not included in the idiolectal set.

course or for payment. All participants were at least 18 years old and were native English speakers with normal hearing.

### 2.1.2. Design

Sixty-four participants were randomly assigned to one of two between-subject exposure conditions in which they performed an auditory lexical decision task. The critical difference between the exposure conditions was whether the ambiguous segment ( $\sim$ sf) was contextually-constrained (dialectal condition) or not (idiolectal condition). In each case, the Exposure phase was followed by a Category Identification task in which participants categorized items on an [asi]-[afi] continuum; the consonants ranged from very /s/-like ([asi]) to very /f/-like ([afi]), with four ambiguous items in between. Listeners were asked to identify each item that they heard as /s/ or /f/. An additional 32 participants served as controls for the [asi]-[afi] category identification task; they only identified members of this continuum, without undergoing the perceptual learning phase.

Seventeen of the experimental subjects in each condition also categorized items on an [astri]-[aftri] continuum. The purpose of this latter continuum was to test learning in a context-specific environment, as learning in the dialectal condition potentially might be. For those participants who were tested on both continua, the order of presentation of the continua was counterbalanced. An additional 17 participants served as controls for the [astri]-[aftri] test, performing only the category identification task, with no previous perceptual learning phase.<sup>4</sup>

To assess the participant's own dialect, each experimental participant also completed a language questionnaire and read aloud ten sentences, which were later analyzed for the pronunciations of our segments of interest ([s], [f], and [s] before [tr]).

### 2.1.3. Materials and procedure

**2.1.3.1. Perceptual learning. Stimulus selection.** Two experimental lists were created for use in the auditory lexical decision task, each with 100 words and 100 nonwords. The idiolectal list contained 20 normally-pronounced /f/ words, 20 critical /s/ words, 60 filler words, and 100 filler nonwords. The dialectal list was identical to the idiolectal list, except for the critical /s/ words. In the idiolectal condition, there were 20 such words, which contained /s/ in varied phonetic contexts (but never before [tr]; e.g., *hallucinate*, *dinosaur*). In the dialectal condition, there were 15 critical words, and these contained /s/ only in contexts where it was followed by [tr] (e.g., *pastry*, *orchestra*). The /s/ in each critical word in both conditions was replaced with an ambiguous fricative midway between /s/ and /f/ ( $\sim$ sf). The dialectal list also included five words that contained /s/ in other contexts (e.g., *pedestal*, *democracy*). These tokens of /s/ were intact (they were *not* replaced with an ambiguous segment),

<sup>4</sup> Our original design for Experiment 1 did not include the [astri]-[aftri] test. When we were halfway through data collection, we decided that having this additional measure might be useful, and we thus had half of the participants label both continua (and we added an equal number of control participants). As will become evident in the Results, this somewhat messy execution nonetheless yielded extremely clear results.

in order to reinforce the dialect: That is, rather than having to infer that /s/ is pronounced as ~sf only in the context of a following [tr], participants were given evidence for the context-specific nature of the change.

The criteria used to select the critical and filler words were the same as in previous experiments (e.g., Kraljic & Samuel, 2005): Words ranged in length from two (*obscene*) to four (*hallucinate*) syllables. Each critical word had a single instance of /s/, and the two sets of critical words were matched to each other and to the filler words in mean number of syllables and mean frequency. The 100 filler words contained no instances of the critical phonemes /s/ or /ʃ/.

We also created a nonword from each filler word. The 100 nonwords contained no /s/ or /ʃ/, and were created by changing several phonemes in each filler word (usually, one phoneme per syllable); phonemes were changed to another phoneme with the same manner of articulation (i.e., glides changed to glides, stops to stops, etc.). To ensure equal numbers of Word and Nonword responses in the lexical decision task, we used all 100 nonwords, 60 of the filler words, 20 /ʃ/ words, and the 20 critical words containing /s/. Appendix A lists all of the critical words, fillers, and nonwords in each experimental group.

*Stimulus construction and procedure.* As in previous experiments (e.g., Kraljic & Samuel, 2005), each critical item was recorded twice by a female speaker: Once with an /s/ (e.g., *hallucinate*), and once with an /ʃ/ in place of the /s/ (e.g., *halluʃinate*). The acoustic properties of /s/ and /ʃ/ enable a fairly straightforward mixing of the two waveforms. /s/ and /ʃ/ are very similar in both duration and amplitude; the main difference between the two is in frequency: /s/ is higher frequency than /ʃ/. Thus, using a sound editing program (Goldwave), the /s/ and /ʃ/ for each item were mixed with various amplitude proportions, until a mixture was created that the authors identified as ambiguous. In this way a unique ambiguous mixture for each critical item was selected for use in the lexical decision (Exposure) phase of the experiment. Subsequent analysis of the mixtures chosen for each set (idiolectal and dialectal) confirmed that the mean frequency of the ambiguous ~sf was matched across the two sets (idiolectal critical items: 5397 Hz; dialectal critical items: 5213 Hz;  $t(34) = 1.56, p = .13$ ).

The continua used in the category identification phase were created in the same way. The same female speaker recorded the syllables [asi] and [afi] (and [astri] and [aftri]). A series of ambiguous ~sf tokens was created by varying the amplitude weighting of /s/ to /ʃ/ in 5% increments (e.g., from 100% /s/ + 0% /ʃ/ to 0% /s/ + 100% /ʃ/). For each continuum, six of the mixtures were selected for use in the experiment based on the authors' judgments and pilot testing.

Participants were randomly assigned to one of three groups: dialectal ~sf, idiolectal ~sf, or control. Those in the control group immediately performed one of the two category identification tasks; they did not perform any prior lexical decision task. Those in the two experimental groups first performed a lexical decision task, followed by the category identification task. Participants were tested individually in a sound shielded booth. Stimuli were presented over headphones and participants responded Word or Nonword (for the lexical decision task) and S or SH (for the categorization task) by pressing the corresponding button on a response panel.

The experimenter stressed the importance of both speed and accuracy on both tasks. Participants were not told that some of the words in the lexical decision task would have ambiguous sounds. Items were presented in a new random order for each participant, with a new item presented 1 s after the participant had responded to the previous item. If a participant failed to respond within 4 s, the next item was presented.

*2.1.3.2. Production (post-test) and questionnaire.* Participants in the experimental conditions completed two final tasks, a production task and a language questionnaire. Both were designed to assess participants' familiarity with and production of the dialectal pronunciation of  $\sim$ sf, and thereby to distinguish participants who produce dialectal  $\sim$ sf from those who do not. Because these tasks potentially could call attention to the nature of the experiment (dialectal variations in pronunciation), we placed them at the end of the experimental session. It is conceivable that having done the perceptual learning task before the production task might make it more difficult to distinguish dialectal from non-dialectal participants, because the range of pronunciations may have narrowed to reflect the exposure. But even with a potentially narrowed range, there was still a strong distinction.

We constructed 10 sentences that collectively included 11 critical segments: 4 /s/; 4 /f/; and 3 /s/ in [str] contexts. Participants were asked to read each sentence aloud into a microphone. After this reading task, they heard a male speaker, with a heavy (natural) Long Island/New York (LI/NY) dialect, speaking each of the ten sentences, and they were asked to imitate the speaker's pronunciation. This speaker was not the (female) speaker participants had heard during exposure. Participants would hear one sentence, repeat it into a microphone (taking care to try to imitate the speaker's pronunciation), and then press a button to hear another sentence. A different random order was used for each participant. The experimenter was not present in the room during the imitation, so that the participants could imitate what they were hearing without feeling awkward. For both the reading and the imitation tasks, each participant's speech was recorded on a digital voice recorder, and later transferred to a computer.

The first purpose of the production task was to evaluate, using the participant's read productions, whether a participant produced the dialectal pronunciation of [str]; i.e., whether the participant pronounced /s/ in this context as /s/ or as  $\sim$ sf. This allows us to assess whether perceptual adjustment to dialectal  $\sim$ sf depends on whether a speaker produces that dialectal variant to some extent already. The questionnaire was designed to provide a subjective correlate to the production task: We asked participants questions about their dialectal background and experience with the New York dialect. See [Appendix B](#) for the questionnaire.

The second purpose of the production task was to evaluate whether our participants *could* produce the  $\sim$ sf pronunciation when explicitly asked to try to imitate speech that contained this segment, independently of whether they do so spontaneously. This point will be elaborated when we discuss how people may or may not accommodate their speech to one another.

## 2.2. Results and discussion

### 2.2.1. Lexical decision

We replaced any participant whose lexical decision accuracy was below 70%. Two of the 64 experimental participants were replaced for this reason. Because the two conditions (idiolectal and dialectal) had different numbers of ambiguous items (20 and 15, respectively), and different contexts in which the ambiguous  $\sim$ sf occurred, performance on the two conditions was analyzed separately. Table 1 presents the lexical decision results.

*2.2.1.1. Dialectal  $\sim$ sf.* Accuracy was very high among participants in the dialectal condition: Participants correctly labeled ambiguous items as words (97.5%) as well as they labeled unambiguous items as words (98.1%),  $F1(1,31) = .33$ ,  $p = .57$ ;  $F2(1,14) = .68$ ,  $p = .42$ . In fact, they were slightly faster to make their judgments for ambiguous items (939 ms, versus 972 ms for unambiguous items),  $F1(1,31) = 6.51$ ,  $p = .02$ ;  $F2(1,14) = 5.345$ ,  $p = .04$ . This difference in speed likely reflects the fact that most of our participants have been exposed to this STR variant quite frequently as part of the LI/NY dialect.

*2.2.1.2. Idiolectal  $\sim$ sf.* Although overall accuracy in the idiolectal condition was high (97.5%), it was slightly but significantly lower for the ambiguous items (95.6%) than for the unambiguous items (99.5%),  $F1(1,31) = 25.8$ ,  $p < .001$ ;  $F2(1,19) = 15.73$ ,  $p = .001$ . In contrast to the dialectal condition, participants in the idiolectal condition were significantly slower to respond correctly to ambiguous items (1043 ms) than they were to respond correctly to unambiguous items (969 ms),  $F1(1,31) = 23.6$ ,  $p < .001$ ;  $F2(1,19) = 10.52$ ,  $p = .004$ .

These data are similar to what we have found in previous experiments using idiolectal variation (e.g., Kraljic & Samuel, 2005). Overall, it seems that while all of the  $\sim$ sf items were perceived as relatively natural, those with  $\sim$ sf in dialectal contexts were less problematic for listeners than those with  $\sim$ sf in other phonetic contexts.

Table 1

Experiment 1, mean accuracy and reaction times (for correct items) for natural and ambiguous critical words, dialectal and idiolectal conditions

	Critical words	
	Natural /f/	Ambiguous $\sim$ sf
<i>Dialectal</i>		
% Correct	98.1%	97.5%
RT (in ms)	972	939
<i>Idiolectal</i>		
% Correct	99.5%	95.6%
RT (in ms)	969	1043

### 2.2.2. Category identification

For each participant, we calculated the average percentage of test syllables identified as /ʃ/. Our critical tests include a comparison of the two experimental groups (dialectal versus idiolectal) to one another, and a comparison of each experimental group to the control condition. To the extent that perceptual learning occurs there will be a lower proportion of SH responses.

**2.2.2.1. [asi]-[afi] continuum.** Participants exposed to idiolectal ~sʃ categorized significantly fewer items as SH than did participants exposed to dialectal ~sʃ ( $F(1, 62) = 10.2, p = .002$ ). A comparison of the learning effect for each group reveals why: For the idiolectal condition, there was a robust perceptual learning effect. Participants who had been exposed to ~sʃ in /s/ words categorized significantly fewer items as /ʃ/ than participants in the control condition, who had no previous exposure (49.4–59.1%, respectively;  $F(1, 62) = 6.51, p = .01$ ). The left panel of Fig. 1 illustrates this effect.

In contrast, exposure to dialectal ~sʃ did not result in perceptual learning (Fig. 1, right panel). Dialectally-exposed participants categorized almost exactly the same percentage of items as /ʃ/ as those in the control condition (58% versus 59%;  $F(1, 62) = 0.065, p = .8$ ).

**2.2.2.2. [astri]-[aftri] continuum.** It is conceivable that dialectal participants did not show perceptual learning on the [asi]-[afi] continuum because learning of the ~sʃ variant might have been restricted to cases in which /s/ is followed immediately by [tr]. Recall that to test this, half of our participants also categorized items on an [astri]-[aftri] continuum. The pattern of results was exactly the same as that for the [asi]-[afi] continuum: Exposure to dialectal ?sʃ did not result in perceptual learning. Dialectal

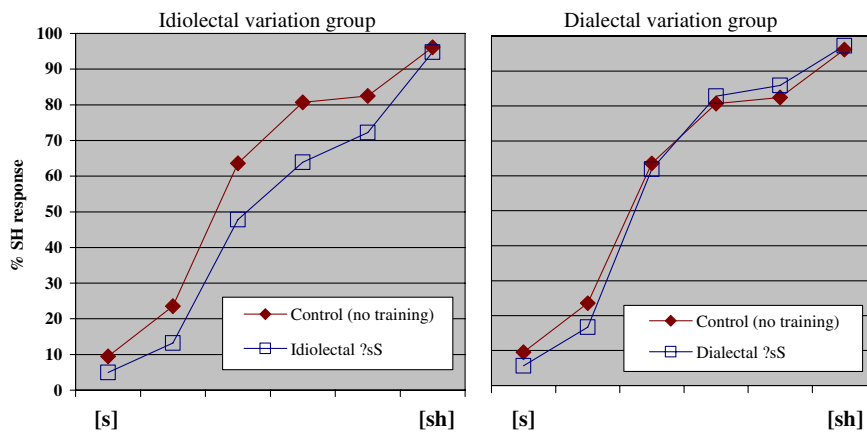


Fig. 1. Experiment 1. Left panel: Participants who trained on /s/ tokens that were ambiguous regardless of the immediate phonetic context showed perceptual learning. Right panel: Participants who trained on /s/ tokens that were ambiguous only when followed by [tr] did not show perceptual learning.

participants categorized essentially the same percentage of items as /j/ as control participants did (51.2% versus 53.9%;  $F(1, 32) = 0.33, p = .34$ ). In contrast, participants in the idiolectal condition showed robust perceptual learning when tested on this continuum, just as they had on the other continuum: They categorized significantly fewer items as /j/ than did control participants (44.5% versus 53.9%;  $F(1, 32) = 5.35, p < .001$ ). This result demonstrates that exposure to an idiolectal pronunciation leads to a general retuning of the phonetic category, even in phonetic contexts that were not among those in the exposure set.

Clearly, then, the context in which an ambiguous segment occurs significantly affects subsequent perception of that segment<sup>5</sup>: Participants who heard ambiguous ~sj in the context of a following [tr] did not show perceptual learning, whereas those who heard the same segment in other contexts did. The question is *why?*

Recall that on the lexical decision task, participants were significantly less accurate and slower to respond to ambiguous items in the idiolectal condition, but *not* in the dialectal condition. These results suggest that the perceptual system handles contextually-driven variation differently than speaker-driven variation. Perhaps the system resolves contextually-driven variation by assigning features of a segment that is ambiguous to the surrounding segments (as Gow, 2003 has argued). Treating such ambiguous /s/ tokens as a form of assimilation could lead to successful interpretation of the phoneme, without any need to change the underlying representation.

### 2.2.3. Production (post-test)

The production data from two participants were lost due to recording error; the data from another two participants had to be discarded because the participants did not produce enough critical tokens (during the imitation task, one participant produced none of the [str] tokens, while the other produced none of the [s] tokens). Recall that each participant produced each of the critical words twice: For the first production, they simply read 10 sentences aloud (Read speech), while for the second production, they heard the 10 sentences produced by a male New York speaker, and were asked to reproduce the sentences while trying to imitate the speaker's pronunciation (imitated speech).

We used the Read speech was to separate our participants into two groups: Participants who naturally manifest the [str] dialect (i.e., /s/ is ~sj-like in [str] contexts) versus participants who do not. We first computed the spectral mean for each participant's pronunciation of /s/ in any context other than [str] (S); of /s/ in [str] context (STR); and of /j/ (SH). (For all speakers, the spectral mean for S is always higher than the spectral mean for SH, and the spectral mean for the /s/ in STR

<sup>5</sup> To ensure that the different results were not due to the different number of ambiguous items in the two groups (the dialectal group had 15 such items, while the Idiosyncratic group had 20), we ran a follow-up experiment in which both groups had only 15 ambiguous /s/ items. The results were identical to what we found here: The dialectal condition did not significantly differ from the control group (58.4–59.1%,  $F(1, 62) = .033, p = .856$ ), and the Idiosyncratic condition was different from these (51.7–58.4%,  $F(1, 62) = 5.102, p = .027$ ). We have also found equivalent perceptual learning using 10 critical tokens versus 20 such tokens, using a /d/-/t/ contrast (Kraljic & Samuel, 2007).

contexts falls either around the S mean, or somewhere between the two.) We next calculated the difference between each participant's S and their STR as a percentage of the difference between their S and SH. Thus, a person's STR production could be assigned a value of 0% (no lower than their S pronunciation), a value of 100% (equal to their SH pronunciation), or any number in between. As an example, a person with an average spectral mean of 8000 Hz for S, 4000 Hz for SH, and 5000 Hz for STR would yield a value of 75%, because that person's pronunciation of /s/ in STR contexts differed from the pronunciation of /s/ in other contexts by 75% of the range between /s/ and /ʃ/; the STR pronunciation was much closer to the pronunciation of SH. This person would therefore be placed in the LI/NY dialect group.

Using this (relatively coarse) criterion resulted in a dialect group with 39 participants (STR proportions ranged from 54% to 100%, with an average of 83%), and a non-dialect group with 21 participants (STR proportions ranged from 0% to 35%, with an average of 15%). The speaker we used in our imitation task (who has a relatively strong LI/NY accent, as judged by the first author and a phonetician), had an STR value of 65%. For analysis purposes, participants with an STR proportion 50% or greater were characterized as LI/NY dialect; those with values below 50% were non-dialect.<sup>6</sup>

Separating participants into two groups on the basis of their natural pronunciations enabled us to address several questions related to the production and perception of a particular variation, as a function of dialect. A preliminary set of analyses showed no effect of participants' dialect on the perceptual shifts reported above (idiolectal ~sj is encoded and represented by the perceptual system, whereas dialectal ~sj appears not to be, regardless of the listener's own dialect). We therefore focus here on possible effects on production.

The central question for production is whether dialect affected performance when participants were asked to imitate our male New York speaker (who has a strong NY accent). The data show an intriguing pattern. Non-dialectal participants changed their productions of STR to be closer to their SH productions by almost 35%; the spectral means of their STR productions were reduced from 5881 Hz to 4862 Hz ( $t(40) = 2.51$ ,  $p = .014$ ). Their STR productions became ambiguous between /s/ and /ʃ/, and became more consistent with the dialect they had just heard than they had previously been. Their production of /s/ in other contexts (S) did not change at all in response to the spoken input ( $t(40) = .02$ ,  $p = .98$ ); only the relationship of STR to S changed. Note that this robust shift occurred even

<sup>6</sup> The acoustic classification did not always match a given participant's belief about whether he or she had a Long Island or New York dialect. In fact, only about half of the participants assigned to the dialect group self-reported having this dialect (18 out of 39 participants); similarly, about half of the participants characterized as non-dialect reported having the dialect (9 out of 21). We relied on the acoustic characterizations, rather than the participants' self-reports, in creating our dialect groups, primarily because the dialectal manifestation we are interested in (movement of /s/ toward /ʃ/ when followed by [tr]) is difficult for listeners to hear (and therefore to self-report). It is also, as we mentioned, possible to have LI/NY dialect and not produce STR, or to produce STR as part of a different dialect.

though during debriefing many participants indicated that they could not hear the variation.

As for the dialectal participants, we divided them into those whose initial S-STR difference was smaller than the difference in the voice they were imitating (i.e., below 65%; their average difference was 57.8%), versus those whose initial S-STR difference was above that of the target voice (i.e., above 65%; their average difference was 88.1%). In each group the STR productions moved in the direction of the voice they were imitating. Those who started out below 65% increased the difference by an average of 7.4%, to 65.2% (even with the small number of subjects in this group, this difference was marginally significant:  $t(10) = 1.88$ ,  $p = .09$ ), and those who started out above 65% *decreased* the difference by 16.3%, to 71.8%,  $t(64) = 2.72$ ,  $p = .008$ . These data show that listeners can, and do, adjust their productions to reflect recently encountered speech when they are asked to imitate the speaker's pronunciation of a given utterance. They even adjust an aspect of their production that they are unaware of: The /s/-[str] shift. In Experiment 2, we look at whether people do so spontaneously; that is, whether their productions of new utterances match their perceptual adjustments.

### 3. Experiment 2

The purpose of Experiment 2 was to replicate Experiment 1's perceptual learning findings, and to see whether perceptual adjustments are mirrored in subsequent spontaneous production. To address these questions, we combined the perceptual learning method with a discourse context in which we hoped to detect any shift in production that might occur.

The perceptual learning method was identical to that of Experiment 1, with the same contrast between a dialectal generator and an idiolectal generator of the odd ~sf pronunciation of /s/. As before, participants then categorized /asi-/afi/ tokens, completed a post-test questionnaire, and read and imitated the 10 sentences.

Critically, both before exposure and again after category identification, participants interacted with the female speaker whose voice they heard during exposure and testing. The female speaker and the subject completed a story aloud (Mad-Libs) using a list of words provided to them. The words included eight tokens of [str] (e.g., *destroy*, *catastrophic*), eight tokens of /s/ in contexts other than [tr] (e.g., *Minnesota*, *rollerskates*), and eight tokens of /ʃ/ (e.g., *ravishing*, *Washington*). We were thus able to see whether participants' productions of these 24 tokens changed in response to the input they received during exposure, and whether any change in production was a function of a change in perception. In order to ensure that the subjects' experience with the female speaker's pronunciations was as controlled as possible, the speaker was *not* the experimenter during this experiment (or in Experiment 1). Thus, the subjects heard the speaker's voice during exposure and test, and interacted with her during the MadLibs, but all instructions and answers to subjects' questions about the experiment were provided by the (male) experimenter.

### 3.1. Method

#### 3.1.1. Participants

Thirty-six undergraduate psychology students from the State University of New York at Stony Brook participated for a research credit in a psychology course or for payment. All participants were 18 years of age or older, and all were native English speakers with normal hearing. None participated in Experiment 1.

#### 3.1.2. Design

The participants performed a story-completion task, followed by a lexical decision task, a category identification test, and a second story-completion task. Finally, each participant completed a post-test questionnaire after reading aloud and imitating 10 sentences.

Participants were randomly assigned to one of two lexical decision conditions: Dialectal  $\sim$ sf or idiolectal  $\sim$ sj. The pre- and post-story tasks, the category identification task, the questionnaire, and the read/imitated sentences were identical for all participants.

#### 3.1.3. Materials and procedure

The materials and procedure for the lexical decision task, the category identification task, the questionnaire, and the read/imitated sentences were identical to those in Experiment 1.

The new task for Experiment 2 was the story-completion task, which participants completed twice: Immediately before the lexical decision task (Before) and immediately after the category identification task (After). The procedure for the task was the same in both the Before and After cases: Participants were given a list of 46 words that appeared in alphabetical order. The words included various parts of speech and functions (e.g., nouns, adjectives, proper names, verbs, etc.). The female speaker with whom participants worked was given two stories, each of which had 23 blank spaces. Each blank space was marked with the appropriate part-of-speech that would fit the space (e.g., *Who could really (verb)\_\_\_\_\_ that there were two (plural noun) \_\_\_\_\_ in space?*) (See [Appendix C](#) for the list of words and [Appendix D](#) for a sample story). The female speaker asked the participant for a word that would fit into each space (e.g., *now I need a verb*), so that by the end of the session, the participant had produced all 46 words. Then the participant was given the two completed stories to read aloud. The female speaker was very careful not to use words with /s/ (in *any* context).

Critically, the 46 words were identical in the Before and After sessions (although the resulting stories were different), providing a within-item test for any changes in the production of [s], [str], and [ʃ] between the Before and After tasks. Each session was digitally recorded. All analyses were based on the participants' natural productions of each word, while they were interacting with the speaker, and *not* on their productions as they read the story aloud. The spectral mean for each of the critical (/s/, [str], and /ʃ/) tokens was obtained for both the Before and After productions of those words.

### 3.2. Results and discussion

#### 3.2.1. Lexical decision

As in Experiment 1, we analyzed performance on the idiolectal and dialectal conditions separately, because the two conditions had different numbers of ambiguous items (20 and 15, respectively), and different contexts in which the ambiguous ~sf occurred.

The results replicate those in Experiment 1 (see Table 2). All participants exceeded the 70% accuracy inclusion criterion on the lexical decision task. Accuracy was very high in both the dialectal condition (99.3%) and in the idiolectal condition (96.2%). As in Experiment 1, the two groups diverged slightly in their labeling of ambiguous versus unambiguous items. Participants in the dialectal condition correctly labeled ambiguous items as words (99.3%) as well as they labeled unambiguous items as words (99.3%),  $F_1 < 1$ ,  $F_2 < 1$ , n.s. Again, they were slightly faster on ambiguous items (1062 ms versus 1101 ms for unambiguous items;  $F_1(1, 17) = 5.882$ ,  $p = .03$ ;  $F_2(1, 14) = 6.282$ ,  $p = .03$ ).

Participants in the idiolectal condition, who heard the same ambiguous segment in different phonetic contexts, were slightly, but not significantly, less accurate at labeling those items as words (94.7%) than at labeling the unambiguous items (97.7%),  $F_1(1, 17) = 2.5$ ,  $p = .13$ ;  $F_2(1, 19) = 3.29$ ,  $p = .09$ . These participants were significantly slower to accurately label ambiguous items (1159 ms) than to accurately label unambiguous items (1067 ms),  $F_1(1, 17) = 8.22$ ,  $p = .01$ ;  $F_2(1, 19) = 9.21$ ,  $p = .007$ . As in Experiment 1, these data suggest that all of the ~sf items were perceived as relatively natural.

#### 3.2.2. Category identification

The findings completely replicate the results of Experiment 1 (see Fig. 2). Participants in the idiolectal condition categorized significantly fewer items as /s/ than did participants in the dialectal condition,  $F(1, 34) = 24.75$ ,  $p < .001$ ; ~sf was learned when it was presented as an idiolect, but not learned when presented as a dialectal variant. Compared to the control condition (from Experiment 1), participants in the idiolectal group categorized significantly fewer items as /s/ (59.1–47.7%,  $F(1, 48) = 6.24$ ,  $p = .02$ ). Participants who were trained on dialectal ?sʃ, on the other

Table 2

Experiment 2, mean accuracy and reaction times (for correct items) for natural and ambiguous critical words, dialectal and idiolectal conditions

	Natural /s/	Ambiguous ~sf
<i>Dialectal</i>		
% Correct	99.3%	99.3%
RT (in ms)	1101	1062
<i>Idiolectal</i>		
% Correct	97.7%	94.7%
RT (in ms)	1067	1159

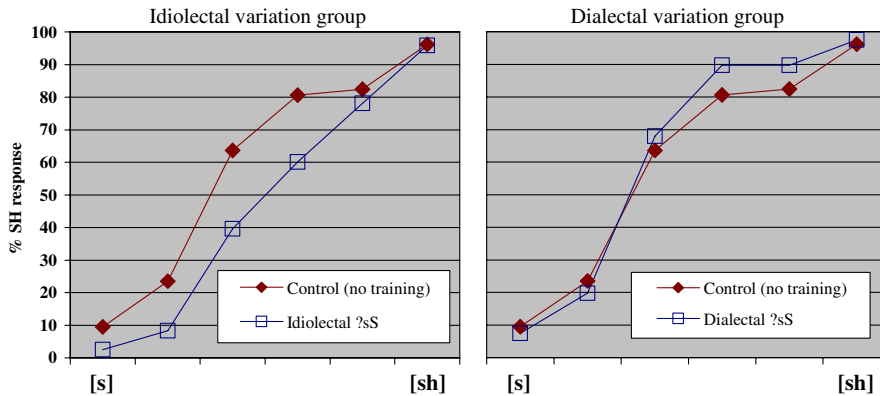


Fig. 2. Experiment 2, Left panel: Participants who trained on /s/ tokens that were ambiguous regardless of the immediate phonetic context showed perceptual learning. Right panel: Participants who trained on /s/ tokens that were only ambiguous when followed by [tr] did not show perceptual learning.

hand, showed no such perceptual learning (61% versus 59.1%),  $F(1,48) = .125$ ,  $p = .7$ .

The findings of Experiments 1 and 2 demonstrate that the context in which an ambiguous segment occurs significantly affects subsequent perception of that segment. When the variation can be attributed to the immediate phonetic context, no changes are made to the representation of that segment. In contrast, exposure to the same ambiguous segment in a phonetic context that cannot account for the variation does cause a representational adjustment: Perceptual learning.

In Experiment 1, we found that participants were generally able to model the target variant even when they were, by their own reports, unaware of it. In Experiment 2, we examined a different aspect of the relationship between perceptual learning and production: When listeners adjust their perceptual representations, what effect, if any, does this have for the subsequent *production* of the critical segment? We know from Experiment 1 that people can change their productions when they are trying to do so. Here, we focus on whether such changes will occur spontaneously as a result of perceptual learning. If production and perception are inherently linked, as in some accounts of language processing (e.g., Pickering & Garrod, 2004), then a change in representation as a result of perceptual experience should result in a corresponding change in the production of that phoneme. On this view, listeners who had been exposed to idiolectal  $\sim$ sf should subsequently change their production of /s/ to be more /ʃ/-like; listeners exposed to dialectal  $\sim$ sf (where no perceptual learning occurs), however, should show no such change in production.

Alternatively, accommodation in production may be flexible – sensitive to pragmatic constraints – and not merely governed by low-level perceptual changes. This possibility is consistent with sociolinguistic accounts in which speakers can modify their pronunciations to either converge or diverge with their interlocutors' (e.g., Bourhis & Giles, 1977; Coupland, 1984). Pragmatically, imitating a speaker's speech

impediment or foreign accent is considered to be rude, whereas converging with an interlocutor's dialectal variation has been proposed as a way to show solidarity or affiliation (see Bourhis & Giles, 1977; Chambers, 1992; Labov, 1973, 1990). If accommodation in production is indeed flexible, participants in the dialectal condition (despite showing no perceptual change) should be more likely to change their production of the critical phoneme than participants in the idiolectal condition (where such a change could be construed as socially inappropriate).

### 3.2.3. Before and after production task (story-completion)

The story-completion task yielded measures of each participant's pronunciation of /s/ in any context other than [str] (S), /s/ in [str] context (STR), and /ʃ/ (SH). We obtained spectral means for each of these cases both Before participants were exposed to any speech, and After they had been exposed to a speaker with either a dialectal variant of /s/, or an idiolectal variant of /s/. If modifications to segmental categories brought about by perceptual learning compel accommodation in production, the STR pronunciation of participants in the dialectal condition should not change, as they showed no perceptual learning. Participants in the idiolectal conditions, in contrast, should shift their pronunciations of S (and perhaps of STR as well). If, on the other hand, social appropriateness constrains accommodation in production, those in the idiolectal condition should not shift their pronunciations following exposure; those in the dialectal condition might (but only if such a shift does not depend on perceptual learning and if the social incentive is sufficiently strong).

We thus made two calculations based on each participant's spectral means. As in Experiment 1, we calculated the difference between each participant's S and STR as a percentage of the difference between his or her S and SH, yielding an STR production value between 0% (equal to the S pronunciation) and 100% (equal to the SH pronunciation). This value reflects any movement of STR pronunciation toward the participant's SH pronunciation. The second calculation was the difference between each person's S and SH, Before and After exposure. If participants change their S productions to reflect the speech they have been exposed to, this difference would be smaller After exposure than Before (note that this change would be compatible with the input for the idiolectal group).

As Table 3 shows, Before and After productions remained remarkably consistent. Due to the random assignment of participants to conditions, there was no initial difference between the dialectal group's pronunciation of S, SH, or STR and the corresponding pronunciations of the idiolectal group (for S:  $t(34) = .32$ ,  $p = .75$ ; for SH:  $t(34) = .32$ ,  $p = .64$ ; for STR:  $t(34) = .49$ ,  $p = .62$ ). Despite subsequently being exposed to different pronunciations of /s/ during the lexical decision task, there was also no difference between the two groups' spontaneous pronunciations After such exposure (for S:  $t(34) = .25$ ,  $p = .79$ ; for SH:  $t(34) = .39$ ,  $p = .69$ ; for STR:  $t(34) = .12$ ,  $p = .9$ ).

The fact that the groups' pronunciations remained so similar to one another before and after exposure reflects the fact that *neither* group's pronunciation, of *any* of the three segments, changed to reflect the speech they had been exposed to.

Table 3

Experiment 2, spectral means for participants' pronunciations of critical phonetic segments (/f/, /s/ in [tr] context, and /s/) before and after lexical decision exposure

	Before	After	Difference (After–Before)
<i>SH</i>			
Dialectal exposure (4516)	3343	3386	43
Idiolectal exposure (4516)	3411	3451	40
<i>STR</i>			
Dialectal exposure (5156)*	4707	4882	175
Idiolectal exposure _N/A	4880	4928	48
<i>S</i>			
Dialectal exposure (6454)	7351	7182	–169
Idiolectal exposure (5385)*	7247	7096	–152

The spectral means for the exposure voice appear in parentheses (ambiguous pronunciations are starred).

Most strikingly, participants in the idiolectal group, who had learned to *perceive* /s/ as more /f/-like, nevertheless did not change their spontaneous production of /s/ to be more /f/-like. The difference between S and SH for the idiolectal group was 3836 Hz Before exposure, and 3644 Hz After exposure ( $t(34) = .53$ ,  $p = .59$ ); the idiolectal group also did not change their production of /s/ in STR contexts to be more like their SH productions ( $t(34) = .03$ ,  $p = .98$ ). Participants exposed to the dialectal variant, where ambiguous /s/ only occurred in [str] contexts, likewise failed to change their production of STR to be more similar to their SH pronunciations: Their STR-SH difference was the same Before and After exposure (1363 Hz versus 1495 Hz;  $t = .32$ ,  $p = .75$ ).

Clearly, exposure to the speaker's voice during perceptual learning did not lead listeners to spontaneously change their speech to become more similar to hers during subsequent interaction. In light of this non-change, it is important to confirm that the participants *could* change their STR productions in response to input when they were explicitly trying to do so.

### 3.2.4. Production post-test

The procedures for measuring each participant's read and imitated pronunciations, for calculating differences in spectral means, and for characterizing each participant as LI/NY dialect or non-dialect, were the same as those used in Experiment 1.

Using these criteria resulted in a LI/NY dialect group with 24 participants (STR values ranged from 50% to 100%, with an average of 80%), and a non-dialect group with 12 participants (STR values ranged from 0% to 48%, with an average of 24%). As in Experiment 1, we found that both the dialect and the non-dialect participants' speech changed in the direction that was consistent with the input when the participants were asked to imitate the male LI/NY speaker. Non-dialectal participants changed their productions of STR to be closer to their SH productions by 18%; ( $t(22) = 2.59$ ,  $p = .02$ ); these participants' STR productions became more

ambiguous between /s/ and /ʃ/, and more consistent with the dialect they were hearing.

Conversely, STR productions for the dialectal group became 6% *less* SH-like, again moving in the direction of the input (although the difference was not significant;  $t(46) = .84, p = .4$ ). Participants are *able* to change their production of /s/ to reflect the speech they are imitating.

One interpretation of the dissociation between the imitation and story-completion results is that hearing someone's speech can induce an immediate but temporary change to participants' production, but not a longer-term change (but cf. Flege, 1987). Of course, the participants' goals were also quite different for the two tasks: In one, their goal was to complete a story; in the other, the goal was to imitate the speech that they heard. One other difference is worth noting. The imitation task required participants to produce the *exact* words they had just been exposed to, as opposed to learning something about the speaker's pronunciation and then applying it spontaneously when generating *new* utterances; only the latter requires a link between a new behavior (i.e., a new motor program) and the underlying representation.

Our results suggest that perceptual changes are not sufficient to compel accommodation in production. The implication is that every output cannot automatically be based on the most recent input (as some have argued: Garrod & Anderson, 1987; Pickering & Garrod, 2004). Rather, the data are consistent with more functional accounts of the link between perception and production. The two systems may serve different goals, and thus require representations that can be distinct (see, e.g., Brennan & Hulstien, 1995; Content, Kearns, & Frauenfelder, 2001; Ferreira, Bailey, & Ferraro, 2002; Samuel, 1991).

#### 4. General discussion

The current experiments were designed to address two questions: (1) does the source of a particular variant in pronunciation affect how it will be handled by the perceptual system? and (2) are perceptual adjustments necessarily mirrored in production? It appears that the answer to the first question is *yes*, and to the second question, *no*.

In Experiment 1, listeners heard the same ambiguous segment ( $\sim$ sf) in one of two contexts. For half of the participants, this pronunciation was experienced as a contextually-driven variant: The ambiguity was a result of the immediately following phonetic context, consistent with a known dialect. For the other half, the same pronunciation was experienced as a non-contextual, idiolectal variant. We found robust perceptual learning for idiolectal  $\sim$ sf; but no shift for the contextually-driven dialectal  $\sim$ sf.

We replicated this contrast in Experiment 2. We also had listeners perform a production task both before and after exposure to  $\sim$ sf, to assess the effect that such exposure might have on subsequent production. Despite a robust perceptual learning effect for idiolectal  $\sim$ sf, there was no corresponding change in production. We also found no change in production for listeners who had been exposed to dialectal  $\sim$ sf. In

contrast, the imitation task demonstrated that listeners *can* adjust their production when copying immediate input, even though they do not generate such changes as a consequence of perceptual learning. Our results support the following conclusions:

#### 4.1. Perception

Most clearly, the perceptual learning data show that the perceptual system does not treat all variations in pronunciation equally; acoustically identical segments are processed differently as a function of their cause. Models of spoken language perception must account for such differences. The present data offer several possible directions for model development, with each direction requiring new empirical tests. One possibility is that our findings reflect (perhaps pragmatically-based) differences between how listeners respond to dialects versus to idiolects; an alternative possibility is that they reflect differences between contextually-driven and non-contextually-driven variations (independently of whether those are dialectal). Yet another possibility is that the differences in learning reflect phonotactic knowledge: [ʃtr] is not a phonotactically legal onset in English, while [s], [ʃ] and [str] are. Perhaps no learning of dialectal ~sʃ is necessary because it can be resolved using purely phonotactic knowledge and therefore would never lead to identification difficulty for the listener (despite being acoustically ambiguous) (see e.g., Gaskell, 2003; Gaskell, Hare, & Marslen-Wilson, 1995 for how listeners might use such distributional knowledge).

If the critical determinant is context, rather than acoustically-independent knowledge like phonotactics or (at a higher level) dialectal status, current speech recognition models might be fairly easily revised by including two stages based purely on the acoustic–phonetic properties of the speech. Specifically, when a listener hears a speech segment that does not quite match any canonical phonetic representation, the perceptual system might first try to disambiguate it by assessing any potential coarticulation it embodies. It might do so, for example, by a method such as feature parsing (Gow, 2003; Gow & Im, 2004; see Gaskell & Marslen-Wilson, 1998, for a different process based on a similar idea), in which each segment is parsed into its features (place and manner of articulation, for example), and any ambiguous features are assigned (if possible) to neighboring phonemes (e.g., *this sounds like an /ʃ/ but it must be an /s/ - it's just taken on some of the place of articulation of the following [tr]*). If feature parsing resolves the conflict (as it would for dialectal ~sʃ, for example), the system accesses the correct representation (/s/) but does not expand it or change it in any way to reflect what it has heard. If the acoustic ambiguity cannot be resolved in this way (as is true for idiolectal ~sʃ), the system then engages perceptual learning processes. These processes use lexical information to categorize the segment, and the underlying phonemic representation is revised to reflect this new pronunciation, which is then applied to subsequent perception of that same segment (by that same speaker, perhaps).

If the critical factor in whether or not perceptual learning occurs is instead whether the pronunciation can be attributed to a dialect, rather than to specific contextual factors, the implications for models of speech perception are potentially more complicated. To many speech perception researchers, it may seem unlikely that linguistic

processing and representation can be directly affected by listeners' beliefs and attributions about the speaker. Yet ongoing research in our lab suggests that such attributions can in fact mediate perceptual processing (Kraljic, Samuel, & Brennan, submitted), for example, causing learning to engage when faced with a property of speech that results from an idiosyncratic speaker characteristics, but compensating in some other way for that same property when it can be attributed to an incidental factor (such as a pen in the speaker's mouth). Several studies synthesizing sociolinguistic and perceptual processes similarly support the idea that speech perception is influenced by listeners' attributions about the speaker (Johnson, Strand, & D'imperio, 1999; Niedzielski, 1999; Strand, 1999); Niedzielski, for example, demonstrated that the same synthesized vowel is perceived differently depending on what listeners believe the speaker's nationality to be. If future work continues to support this idea, models of speech perception will have to address a set of issues that they currently do not, such as how the system integrates information about awareness, attention, and attributions (see Baker et al., 2006; see Arnold, Hudson Kam, & Tanenhaus, in press, for the same issues in language comprehension more generally).

#### 4.2. Production

In at least one domain, our results show that awareness is not a critical factor. The imitation data show that changes in production do not depend on awareness of particular pronunciations: People are able to produce variations that they do not report hearing. Our story-completion data also demonstrate a decoupling between changes in production and changes in perception: Despite having shown robust perceptual learning effects, participants (in the idiolectal condition) did not subsequently change their production of the corresponding phonemes. While such perceptual changes might prove to be *necessary* for production changes, they do not seem to be sufficient.

Our results showing production changes (via imitation) and non-changes (despite perceptual shifts) are consistent with previous observations of substantial long-term shifts in perception that are not subsequently mirrored in production (e.g., Cooper, Ebert, & Cole, 1976; Cooper & Nager, 1975; Samuel, 1979). This pattern suggests that the representations underlying perception and production are separate. Yet, people *do* change their production based on variants they have heard, both in the lab (when imitating; and sometimes even when they are not trying to imitate; see Goldinger, 1998) and out of it (speech accommodation reported by sociolinguists). Two ideas may prove useful in future attempts to reconcile these facts: (1) perception/production and social appropriateness, and (2) perception/production and representational levels.

In the Introduction, we discussed the potential effect of attribution: Listeners' perceptual systems appear to be sensitive to the source of variations in pronunciation. Perhaps speech production is sensitive to the same constraints, but *also* relies on perceptual representations. That is, if a perceptual adjustment has been made, the production system will make use of that adjustment, *unless* the adjustment would be inconsistent with pragmatic goals (as in the case of our idiolectal variation). But if a perceptual adjustment has not been made (as in our dialectal

condition), then the production system has no basis for adjustment, so it also does not change. Testing this hypothesis requires a pronunciation that can cause a perceptual change, and for which a production change would be goal-consistent.

There is an alternative potential explanation for the lack of observable changes in production despite such changes in perception: Perhaps well-practiced motor programs mask any potential changes. To understand the implications that such a motor practice filter might have, it is necessary to consider the kinds of representations that are commonly thought to underlie speech production. According to most theories, once a word has been selected for production, the system is thought to generate an abstract phonological code; this code in turn is filled in with phonetic codes, which ultimately get articulated using motor programs (Cholin, Levelt, & Schiller, 2006; Laganaro & Alario, 2006; Levelt, Roelofs, & Meyer, 1999; MacNeilage, 1970; Sevald, Dell, & Cole, 1995). Although any of these levels of representation could be shared by production and perception the motor level has the additional issue of practice to consider. The claim of a hypothesis based on motor practice is that any changes to one or both of the first two levels of representation might be difficult to detect because the final, motor stage is a very slow learner. One way to test this hypothesis would use variations in speech that already have very practiced motor programs (perhaps by recruiting participants who are bi-dialectal).

In any event, our imitation data show that the participants can immediately recruit motor programs that are sufficient to approximate the critical segment when they intend to do so. It may be that the very nature of imitation is to enhance the linkage of perception and production representations, in ways that do not otherwise occur. One form of link may be articulatory in nature, and is posited in the motor theory of speech perception (e.g., see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; for a recent review, see Galantucci, Fowler, & Turvey, 2006). In the motor theory, perception can provide ‘instructions’ for subsequent production, so that speech that immediately follows perception can quite easily be imitative (again, see Galantucci et al., 2006). What our data (and Goldinger’s delayed production data) suggest is that, at least with respect to coarticulation, such articulatory information could be the most immediate object of perception, but it is not preserved (but again, cf. Flege, 1987). Once a percept has faded, what remains is a more abstract code and the speaker’s own practiced motor patterns. Thus, pure auditory perceptual learning can, at most, provide information to the production system about motoric or articulatory goals. The production system must then determine how best to implement those goals, and the motor system must practice producing them before corresponding changes in production can be evident.

Considering the different levels of representation and their role in speech production raises an interesting question: Where are dialectal differences generated in the standard production model? If dialectal differences emerge purely at the motor-pattern level, they should presumably take substantial overtraining to change. Practice articulating the relevant motor patterns should suffice for such articulatory training.

An alternative is that dialectal differences are generated at the phonological or the phonetic level, or in the mapping between them. If so, they could be more malleable, but they should also require exposure to meaningful speech, rather than nonwords or random sequences of syllables. In fact, Norris et al. (2003) demonstrated that the same variation that produced large perceptual learning shifts in known words produced no perceptual learning when the variation was presented in the context of nonwords.

The picture that emerges from the pattern of results that we have found is that not all sources of variation in pronunciation are treated equally by the perceptual system; the differing perceptual processing, in turn, can have different effects on speech production, either because of pragmatic constraints on language use or because of the architecture of the language processing system itself. Given the increasing frequency of contact with people who speak with different accents and different dialects, a better understanding of different types of variation will be increasingly important for spoken language research and for developing successful automated speech recognition systems.

### Acknowledgements

This material is based upon work supported by a National Science Foundation Graduate Fellowship and by NSF Grant No. 0325188 and NIH Grant R0151663. We would like to thank Donna Kat, Carol Fowler, and two anonymous reviewers for their careful and extremely helpful comments.

### Appendix A

Experiment 1, critical and filler words (with nonwords in parentheses) for dialectal and idiolectal conditions, Lexical Decision task

Critical (unambiguous)/f/ words (SH)	Critical /s/ words (S replaced by ~sf)	
	Group1	Group2
<i>Both groups</i>	<i>Dialectal variation</i>	<i>Idiolectal variation</i>
ambition	Abstract	Arkansas
beneficial	Administrator	eraser
brochure	Artistry	coliseum
commercial	Astronaut	compensate
crucial	Australia	democracy
efficient	Chemistry	dinosaur
flourishing	democracy <sup>a</sup>	embassy
glacier	Demonstrate	peninsula
graduation	Destroy	episode

**Appendix A** (*continued*)

Critical (unambiguous)/f/ words (SH)	Critical /s/ words (S replaced by ~sf)	
	Group1	Group2
impatient	hallucinate <sup>a</sup>	hallucinate
initial	Hamstring	legacy
machinery	Illustrate	literacy
negotiate	Industry	medicine
official	Instrument	obscene
parachute	obscene <sup>a</sup>	personal
pediatrician	Orchestra	parasite
publisher	Pastry	pregnancy
reassure	Pedestrian	reconcile
refreshing	personal <sup>a</sup>	rehearsal
vacation	Tennessee <sup>a</sup>	Tennessee

<sup>a</sup> These five words contained normal /s/, to fully represent the dialect (e.g., *the ambiguous segment occurs only immediately preceding a [tr]*).

---

**Filler words (nonwords)**


---

Accordion (igoldion)	Napkin (mibgem)
America (anolipa)	Negate (mikid)
Annoying (imoyem)	Outnumber (admunker)
Armadillo (alnadiro)	Panicky (bimikay)
Bakery (pakelo)	Parable (baliber)
Ballerina (galliwinou)	Parakeet (bawaseet)
Blueberry (pluepelai)	Pineapple (bimobel)
Bullying (pourilar)	Platonic (kradomet)
Camera (ganla)	Remedial (lenediaw)
Crocodile (klogodar)	Romantic (wonontic)
Darken (perkum)	Tactical (dadigal)
Directory (tilegkalo)	Titanium (bikanian)
Document (pogunemd)	Turbulent (durkuwomt)
Domineering (konimeelum)	Tutorial (datiliar)
Dynamite (tymolipe)	Umbrella (omplero)
Embody (enpaiki)	Warrantee (rawamtee)
Gardenia (kaldemia)	Wealthy (lirthy)
Grammatical (kloumidiger)	Withdrawal (rikmaral)
Gullible (kuradel)	Wrinkle (lindel)
Hamburger (hintarber)	<i>Additional Nonwords (original words not used)</i>
Honeymoon (hominaim)	Acominig (economic) l
Hurdle (hilder)	Aigi (eighty)

(continued on next page)

**Appendix A** (*continued*)

## Filler words (nonwords)

---

Identical (itempider)	Ailounam (ornament)
Ignite (aknid)	Amalar (immoral)
Immoral (irimel)	Anemer (enamel)
Inhabit (emhoutic)	Bamtel (panther)
Knowingly (mowery)	Bliparg (predict)
Laminate (wonimtic)	Gairelom (tailoring)
Legally (weekary)	Galliwinou (ballerina)
Liability (riakirity)	Gerbualo (purgatory)
Lobbying (woppakin)	Geypalg (keyboard)
Lunatic (rumatik)	Gilday (tardy)
Lyrical (ryligal)	Gondimually (continually)
Manually (namuery)	Gonedial (comedian)
Marina (nawinow)	Halomimoc (illuminate)
Melancholy (neramgory)	Hiliun (heroine)
Membrane (nempring)	Ibirak (epilogue)
Memory (nomeray)	Imdalier (interior)
Metrical (nekridal)	Ithomel (ethanol)
Military (niritaly)	Kegimel (beginner)
Momentary (nometmoly)	Kelabidel (therapeutic)
	Kermimer (terminal)
	Kerkrun (pilgrim)
	Lilgrai (worldly)
	Logelai (rubbery)
	Loubel (wrapper)
	Maidnow (nightmare)
	Marody (melody)
	Omperog (interact)
	Pirugalo (burglary)
	Rakil (lapel)
	Rengimer (lengthen)
	Rimkuwar (lingering)
	Tamical (cannibal)
	Tonamlemp (commandment)
	Umikory (unitary)
	Ungelnin (undermine)
	Waiper (lethal)
	Wojalto (royalty)
	Joungel (younger)

---

**Appendix B**

Experiments 1, post-experiment language questionnaire:

1. How old are you?
2. Where are you from originally (specify city, state)?
3. How long have you been living on Long Island? (If you do not/have never lived on LI, where do you live and how long have you been living there?)
4. Have you lived in other places besides Long Island? If so, where and for how long (approximately)?
5. Do you know more than one language well enough to follow a conversation? If so, which language(s), and how did you learn them?
6. Can you identify/hear Long Island or New York accents on other people?
7. What are some of the things you hear most when you hear a Long Island or New York accent?
8. Do you think you have a Long Island or New York accent?
9. Do you have friends or family members who have a Long Island or New York accent?

**Appendix C**

Experiment 2, list of words given to each participant for use in the story-completion task

ankle	climbed	George Bush	Mother Teresa	Slug
assassin	destructive	Guatemala	Nostradamus	Strange
baker	earlobe	guitar	Nostril	Super
bathroom	feet	holler	rollerskates	Sushi
Britney Spears	flounder	kangaroo	Sean “P.Diddy” Combs	Toolbox
catastrophic	fly	kayak	shooting	Toothbrushes
cereal	frustrated	Madonna	ridiculous	Washington
claustrophobia	garbage	Minnesota	shoelaces	Wonderific
abolish	illustrate	moon		

**Appendix D**

Experiment 2, sample story that each participant and the female speaker completed together.

## A BEDTIME STORY

Once upon a time, there was a very curious girl named \_\_\_\_\_  
 (Female Name)  
 that was always poking her nose into everybody's \_\_\_\_\_.  
 (plural noun)  
 She kept company with a/an \_\_\_\_\_ man named \_\_\_\_\_.  
 (adjective) (Male Name)  
 who enjoyed buying her many \_\_\_\_\_ presents. Once he gave her a  
 (adjective)  
 diamond \_\_\_\_\_ to wear on her \_\_\_\_\_.  
 (noun) (part of body)  
 and a \_\_\_\_\_ as a pet. He was always taking her traveling, to different  
 (water animal)  
 and \_\_\_\_\_ places. One day they might go to \_\_\_\_\_.  
 (adjective) (location)  
 and the next week they could be in \_\_\_\_\_. But her favorite gift was  
 (location)  
 the day he presented her with a \_\_\_\_\_. As soon as she saw it, she  
 (land animal)  
 proceeded to examine the \_\_\_\_\_ animal closely, for one can  
 (adjective)  
 never \_\_\_\_\_ too carefully. First, she looked at its  
 (verb)  
 \_\_\_\_\_ and then at its \_\_\_\_\_. She even  
 (body part) (noun)  
 opened its mouth so she could get a good look at its \_\_\_\_\_. She  
 (noun)  
 finally decided to name it \_\_\_\_\_. Despite her careful inspection, the  
 (Name)  
 \_\_\_\_\_ suddenly became \_\_\_\_\_ and bit off her  
 (same animal) (emotion/verb ending in ED)  
 \_\_\_\_\_. Her \_\_\_\_\_ was never the same.  
 (body part) (noun)  
 The moral of this \_\_\_\_\_ bedtime story is to never look a gift  
 (adjective)  
 \_\_\_\_\_ in the mouth.  
 (same animal)

## References

- Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (in press). If you say thee uh – you're describing something hard: The on-line attribution of disfluency during reference comprehension. *The Journal of Experimental Psychology: Learning, Memory and Cognition*.
- Baker, A., Mielke, J., & Archangeli, D. (2006). Top-down and bottom-up influences in English S-retraction. Conference presentation, *NWAV 35, New Ways of Analyzing Variation*, November 9–12, Columbus, OH. (<<http://www.u.arizona.edu/~tabaker/NWAV35str.pdf>>).
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13, 145–204.
- Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity, and intergroup relations* (pp. 119–135). London: Academic Press.
- Brennan, S. E., & Hulstien, E. A. (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, 8, 143–151.
- Chambers, J. K. (1992). Dialect acquisition. *Language*, 68(4), 673–705.
- Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205–235.

Please cite this article in press as: Kraljic, T. et al., Accommodating variation: Dialects, idiolects, ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.07.013

- Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language, 45*, 177–199.
- Cooper, W. E., Ebert, R. R., & Cole, R. A. (1976). Speech perception and production of consonant cluster [st]. *Journal of Experimental Psychology: Human Perception and Performance, 2*(1), 105–114.
- Cooper, W. E., & Nager, R. M. (1975). Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model. *Journal of the Acoustical Society of America, 58*(1), 256–265.
- Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Society of Language, 46*, 49–70.
- Coupland, N., & Giles, H. (1988). Communicative accommodation: Recent developments. *Language and Communication, 8*(3 & 4), 175–327.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47–65.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review, 13*(3), 361–377.
- Gallois, C., & Callan, V. J. (1988). Communication accommodation and the prototypical speaker: Predicting evaluations of status and solidarity. *Language and Communication, 8*, 271–283.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.
- Gaskell, M. G. (2003). Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463.
- Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science, 19*, 407–439.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380–396.
- Giles, H. (1973a). Communicative effectiveness as a function of accented speech. *Speech Monographs, 40*, 330–331.
- Giles, H. (1973b). Accent mobility: A model and some data. *Anthropological Linguistics, 15*, 87–105.
- Giles, H., & Powesland, P. F. (1975). *Speech styles and social evaluation*. London: Academic Press (Reprinted as Accommodation theory. (1997). In N. Coupland & A. Jaworski (Eds.), *A Sociolinguistics Reader*, (pp. 232–239). Basingstoke, England: Macmillan.).
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.
- Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory and Cognition, 27*(2), 328–338.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics, 65*, 575–590.
- Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296.
- Johnson, K., Strand, E. A., & D'imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 24*(4), 359–384.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141–178.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*(1), 1–15.
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (submitted). First impressions and last resorts: How listeners adjust to speaker variability.
- Labov, W. (1973). The social setting of linguistic change. In T. A. Sebeok (Ed.), *Current trends in linguistics, 11*, pp. 195–251.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–53). Englewood Cliffs, NJ: Prentice-Hall.

- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205–254.
- Laganaro, M., & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, 55, 178–196.
- Lawrence, W. P. (2000). /str/→/ʃ tr/: Assimilation at a distance? *American Speech*, 75(1), 82–87.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29(1), 117–123.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- MacNeilage, P. F. (1970). Motor control of the serial ordering of speech. *Psychological Review*, 77, 182–196.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.
- Nygaard, L. C., & Pisoni, D. B. (1995). Talker-specific learning in speech perception. *Perception and Psychophysics*, 60, 355–376.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309–328.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24, 175–184.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Brain and Behavior Sciences*, 27(2), 169–225.
- Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication*, 13, 109–125.
- Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651–666.
- Samuel, A. G. (1979). Speech is specialized, not special. Unpublished doctoral dissertation.
- Samuel, A. G. (1991). Perceptual degradation due to signal alternation: Implications for auditory pattern processing. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 392–403.
- Sancier, M., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.
- Sevold, C., Dell, G., & Cole, J. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34, 807–820.
- Shapiro, M. (1995). A case of distant assimilation: str/ →/ʃtr/s. *American Speech*, 70(1), 101–107.
- Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes*. Hillsdale, NJ: Lawrence Erlbaum.
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1), 86–99.
- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, 2, 339–363.