

Perceptual learning, talker specificity, and sound change

MEREDITH TAMMINGA

University of Pennsylvania

WEI LAI

University of Pennsylvania

ROBERT WILDER

University of Pennsylvania

LACEY WADE

University of Pennsylvania

Abstract

Perceptual learning is when listeners hear novel speech input and shift their subsequent perceptual behavior. In this paper we consider the relationship between sound change and perceptual learning. We spell out the connections we see between perceptual learning and different approaches to sound change and explain how a deeper empirical understanding of the properties of perceptual learning might benefit sound change models. We propose that questions about when listeners generalize their perceptual learning to new talkers might be of particular interest to theories of sound change. We review the relevant literature, noting that studies of perceptual learning generalization across talkers of the same gender are lacking. Finally, we present new experimental data aimed at filling that gap by comparing cross-talker generalization of fricative boundary perceptual learning in same-gender and different-gender pairs. We find that listeners are much more likely to generalize what they have learned across same-gender pairs, even when the different-gender pairs have more similar fricatives. We discuss implications for sound change.

1 Introduction

It is widely accepted that in speech perception, listeners have a relatively high degree of perceptual flexibility (Repp & Liberman 1984): they can use a range of contextual cues to dynamically adjust their interpretation of the phonetic input they encounter. In historical phonology, perceptual flexibility is often thought to have an important role to play in sound change because it can both introduce and be a response to the variability that change requires. For example, Ohala's influential model of sound

change (e.g., Ohala 1981) centrally involves compensation for coarticulation, a form of perceptual adaptation to the articulatory context in which sounds occur. Accordingly, the sound change literature has long acknowledged a need to understand the mechanisms of perceptual flexibility in human speech perception. Ohala (1981), to continue the example, points to experimental studies of compensation for coarticulation as an important source of evidence for the basic perceptual mechanisms underlying his theoretical proposal.

It is also understood that perceptual flexibility in the moment is insufficient for sound change. In order for any perceptual adjustment to play a role in longer-term change, listeners must also sometimes allow those adjustments to influence their *future* behavior, in both perception (our focus here) and production. On the perception side, the longer-term influence needs to extend to contexts that go beyond the original perceptual experience—different words and sentences, different social settings, different interlocutors. In other words, listeners need to not only retain but also *generalize* what they have learned. In addition to understanding the mechanisms of momentary perceptual flexibility, then, theorists of sound change need to understand the learning and behavioral mechanisms by which listeners update their expectations about future input. Such mechanisms are targeted by a collection of experimental paradigms eliciting **perceptual learning** (for an overview, see Samuel & Kraljic 2009). In this paper we expand on the connection between perceptual learning and sound change, with special attention given to the question of how listeners generalize across different talkers.

Of course, it is also clear that listeners must go beyond perceptual learning and generalization for sound change to take hold at a community level: they must ultimately integrate what they have heard into their own production targets. Experimental work eliciting phonetic convergence investigates how people incorporate what they perceive into their own speech production behavior in the short term (Shockley et al. 2004; Nielsen 2011; Babel et al. 2014; Zellou et al. 2016), while studies of convergence between people who interact regularly over time have addressed questions about longer-term convergence (Pardo et al. 2012; Sonderegger et al. 2017). While we do discuss questions about production at some length in this paper, that discussion focuses on how perceptual flexibility might underlie short- or long-term adjustments in production; we do not provide an exhaustive account of either the factors that go into convergence or the relationship between convergence and sound change. While we believe such questions are equally important in the bigger picture, the sound change literature is already well connected to experi-

mental and corpus-based research on convergence at different scales. It is our informal observation that historical phonology's connection to ongoing experimental research in perceptual learning is not as well developed, perhaps because the latter has been reported primarily (though not exclusively) in psychology journals. In this paper, we aim to help foster a stronger connection in this regard by discussing how current issues in perceptual learning are relevant to theories of sound change.

This paper has three aims of approximately equal weight. The first, in Section 2, is to make explicit the connections we see between perceptual flexibility and models of sound change. The second, in Section 3, is to synthesize for a historical phonology audience the empirical evidence around talker specificity in perceptual learning. The third, in Section 4, is to report novel experimental data on perceptual learning with possible implications for the role of talker specificity in perceptual learning and sound change. Before we turn to these goals, though, we provide a brief description of what we mean by perceptual learning and how it is elicited experimentally.

1.1 Experimental perceptual learning

Perceptual learning, in the domain of speech perception, refers to the process where a listener (subconsciously) adjusts some perceptual mapping, at least temporarily, in a way that persists *after* listening to novel speech input. Perceptual learning may take the general form of an improvement in comprehension when listening to speech in an unfamiliar accent (Clarke & Garrett 2004; Clopper & Pisoni 2004; Bradlow & Bent 2008). It can also be seen more narrowly when a listener adjusts their perceptual category boundary between two phonemes after listening to a speaker who produces an atypical version of one of those phonemes (Norris et al. 2003; Eisner & McQueen 2005; Kraljic & Samuel 2005). We will mostly discuss studies of the latter type of perceptual learning, which Samuel & Kraljic (2009) term *phonetic retuning*, because in our view it bears the closest relationship to the adoption of a single sound change within a language. Note that the term "perceptual learning" is also sometimes used metonymically to refer to experimental paradigms that elicit the learning behavior, or to the particular response patterns in such experiments, but we follow Samuel & Kraljic in using the term to refer to any learning behavior where exposure to "speech that is in some way noncanonical...produces a change in subsequent language processing" (2009:1208), which we can be elicited through a range of different ex-

perimental methods and also is assumed to occur naturally in everyday language experiences.

Much of the literature we will review in Section 3, as well as our own experimental data that we will report in Section 4, involves a method introduced by Norris et al. (2003) that induces perceptual learning using a training phase in which lexical cues suggest the categorical interpretation of an ambiguous sound to the learner. This “lexically-guided” perceptual learning takes advantage of the well-established phenomenon where listeners prefer to hear words whenever possible (Ganong 1980). For example, if a listener hears [ɟə.ɪæ?], where [?] represents a fricative halfway between [f] and [s], the listener will be inclined to interpret the fricative as /f/ to form the word “giraffe,” because hearing the fricative as /s/ would produce only the nonword “girasse.” Hearing the same ambiguous fricative in [hə.ɪæ?], on the other hand, might lead the listener to an /s/ interpretation. In Norris et al. (2003), participants were randomly assigned to either an /f/-biased or /s/-biased condition (with Dutch stimuli) and were trained on a lexical decision task that consistently signaled the phonemic identity of the ambiguous fricative according to whichever condition they were in. After this training, participants were tested on categorization of the ambiguous fricative in the syllables [ɛf] and [ɛs]. Listeners who had been trained on /s/-biased stimuli were more likely to categorize [?] as /s/ in these syllables than those who had been trained on /f/-biased stimuli, suggesting that the training had shifted either or both groups’ perceptual boundaries between /f/ and /s/. The fact that this and subsequent perceptual learning experiments assess the perceptual shift *after* the training phase (as opposed to in the moment of biased perception) and in underinformative stimuli (as opposed to with the bias-inducing cues present) is useful from a historical phonology perspective, because it shows that phonetic flexibility can “stick.”

Beyond Norris et al.’s lexically-guided paradigm, retuning of a phoneme boundary through perceptual learning can also be guided by a range of other contextual cues, including audio-visual integrity (Bertelson et al. 2003; van Linden & Vroomen 2007; Vroomen et al. 2004, 2007; Vroomen & Baart 2009; van der Zande et al. 2014; Jesse & Kaplan 2019), phonotactic regularities (Cutler et al. 2008), coarticulatory patterns (Connine & Darnieder 2009), and semantic predictability (McAuliffe 2015). We focus on lexically-guided perceptual learning experimental approaches because lexical context is a robust and realistic source of information about the phonemic identity of different phonetic inputs, suggesting that experimental work in this domain can reasonably be thought of as offering a

window into the processes that might be involved in real-world sound changes.

2 The relationship between sound change and perceptual learning

As we have already noted, the idea that experimental evidence on phonetic flexibility of various kinds is relevant to models of sound change is far from novel. Nonetheless, we believe that it is worth making explicit the connections we see between these bodies of work in order to make the motivation for this paper clear. We begin in Section 2.1 by discussing the basic role that perceptual and productive flexibility, enduring over time, play in a range of different approaches to sound change. In Section 2.2, we discuss how more detailed empirical facts about the learning processes involved in both perception and production might contribute to shaping models of change.

2.1 The basic role of perceptual learning in sound change models

A fundamental idea shared by many, perhaps all, models of sound change is that sound change advances when people hear phonetic input from other people around them and end up speaking differently in response. Models of sound change have differed in terms of how much explanatory weight they put on the listener role or the speaker role within this basic picture. While the connection between perceptual learning and listener-oriented models in particular may seem more obvious, we see a role for perceptual learning across many different sound change models. Models of change also differ in whether they primarily aim to solve the actuation problem, of when and why changes are innovated, or are directed at the question of how actuated change spreads through a community. We suspect that perceptual learning might be thought of as involved in both of these stages of change, or perhaps even as bridging them, but we do not attempt to further elucidate this point here. Instead, we survey several previous discussions of sound change to illustrate the basic role that a perceptual learning mechanism must play at some stage in each.

We have already mentioned the Ohalan model of sound change, in which under- or over-compensation can give rise to a “mini-sound change” (Ohala 1981:184) through the component of the model that Ohala terms the “listener-turned-speaker” (Ohala 1981:183). Ohala (1993) spells out in more detail what it takes for a listener to turn into a speaker: “Such variation becomes fossilized if and when listeners fail to recognize the variation as totally predictable from context, incorporate it

into their own mental lexicons, and base their own pronunciation on the new norm” (1993:163). While the exact contours of the Ohalan model in terms of its historical-typological coverage are dependent on details about when particular kinds of errors are likely to arise, it is clear from this quote that a basic perceptual learning mechanism (that is, the mechanism by which listeners “incorporate it into their own mental lexicons”) is required to promote a temporary mental misparsing into a behavioral change that might be adopted by others. While we might be in the habit of thinking about perceptual learning in terms of retuning our perceptual mappings to achieve a closer veridical match to the input, in fact the same kind of perceptual learning process is necessary even if what is being learned is an innovation or “error” relative to the speaker’s intentions.

Subsequent work building on Ohala’s foundations inherits the importance of this listener-turned-speaker component, which as we have just suggested entails a perceptual learning mechanism. For example, Beddor (2009) proposes that the information carried by coarticulatory cues may be interpreted differently by different listeners—and therefore incorporated differently into their production norms downstream. While her framework differs from the Ohalan model in that it does not construe mismatches in the phonetics-phonology mapping as errors *per se*, the listener-turned-speaker’s role in the model is the same. Baker et al., aiming to restrict the Ohalan model, suggest that naturally-occurring but phonetically-extreme instances of coarticulation that happen to be produced by socially influential speakers create the potential for change, positing that, “Given the appropriate social conditions, this potential can be realized through another speaker adopting the novel target in his/her speech” (2011:351). Schertz & Clare, surveying proposals to think of sound change in terms of cue re-weighting or misattribution, summarize the point neatly: “Sound change occurs if and when these perception patterns transfer to production” (2019:5–6). Such transfer is simply another way of describing the listener-turned-speaker, and therefore depends to some degree on perceptual learning.

The models we have discussed so far, sometimes described as “listener-oriented,” are focused on the details of perception. An example of a class of models that give more attention to speaker behavior are those that Auer & Hinskens (2005) call “change by accommodation” models. In change-by-accommodation, convergence between speakers during conversational interactions accumulates over time into community-level change. For example, when Bloomfield introduces his principle of density (of communication) as a major factor in language change, he writes that “every speaker is constantly adapting his speech-habits to

those of his interlocutors; he gives up forms he has been using, adopts new ones, and perhaps oftenest of all, changes the frequency of speech-forms without abandoning any old ones or accepting any that are really new to him" (1933:476). Other examples of change by accommodation models, or discussions thereof, include Paul (1880), Trudgill (1986), Niedzielski & Giles (1996), and Sonderegger et al. (2017).

While we may tend to think of convergence as a speaker behavior, the process of accommodation or convergence to an interlocutor presumably requires that the speaker convert perceptual input into a new (at least temporary) production target. As Auer and Hinskens point out, this step of the model is not fully understood: "there is some ambiguity in the model concerning the driving forces behind the first step, or short-term accommodation" (Auer & Hinskens 2005:337). This ambiguity is partly due to questions about the mechanisms and motivations by which listeners maintain some short-term implicit memory of the phonetic input they encounter: questions, that is, about perceptual learning. Auer and Hinskens also point out that there are many unanswered questions about the circumstances under which interpersonal accommodation in interaction actually leads to longer term production adjustments and ultimately to permanent change in an entire community's language (see also Sonderegger et al. 2017). We would suggest that the properties of perceptual learning's generalization and durability might inform these questions about how and when temporary shifts are cemented into longer-term change.

So far we have suggested that some mechanism(s) by which listeners retain perceptual input and integrate it into their mental lexicon are necessary to produce a change in behavior over time. Discussions of sound change have differed in the extent to which these mechanisms seem not only necessary but potentially sufficient, or nearly so. To what extent does perceived phonetic input, when maintained in memory, feed directly into novel production behavior? Consider, for example, this passage from Martinet (1952):

For each [phoneme]...there must be an optimum which we might call the center of gravity of every range of dispersion, but actual performances will normally fall somewhat off the mark. In the normal practice of speech, some of them are even likely to fall very far off it. If too dangerously near the center of gravity of some other phoneme, they may be corrected, and, in any case, will not be imitated. If unusually aberrant, slightly beyond the normal range of dispersion, but not in a direction where misunderstanding might arise...they might well end up as establishing a legitimate extension of the acceptable range. We shall reckon with a sound shift as soon as the normal range of a phoneme...is being ever so little displaced in one direction or another... (Martinet 1952:4-5)

While the shift outcome is framed in terms of the abstract phonemic unit, the mention of imitation and the broader context of the paper make it clear that Martinet intends for the displacement of the “normal range of a phoneme” to include a displacement in speakers’ production targets. In other words, Martinet assumes that what is perceived, modulo “corrections” to avoid misunderstanding, is converted into what is produced. The computation of a target vowel quality from a pool of perceptual observations shows up again in Labov’s (1994) discussion of functional views such as Martinet’s, although in this case he is arguing against the functional explanation put forward by Martinet for why some production tokens might not be entered into such a computation:

The fronted token of /o/ will no longer be heard as within the range of the /æ/ distribution, and there is a much greater likelihood that it will be identified correctly as /o/. It will then contribute to the computation of the mean value of /o/, and accordingly, that mean will be shifted toward the front...no matter how small the effect, repeated misunderstandings will have the effect of facilitating the shift.” (Labov 1994:587)

We interpret Labov’s reference to the shift that is facilitated to be a shift observed in speech production. In other words, some portion of the perceptual input is maintained in memory and converted into a new production target. The way Martinet and Labov differ here is in their view of the motivation or mechanism by which *some* input is maintained and integrated into future behavior, while other input does not have such an influence. However, both see a role for the retention of perceptual input that we think could reasonably be understood as reflecting perceptual learning.

These particular proposals from Martinet and Labov have much in common with usage-based models of sound change (Bybee 2002; Johnson 2007; Hay et al. 2015; Hay & Foulkes 2016) — in particular, they share a mechanism in which a production target is computed from a pool of input instances. In fact this computation, albeit over a richer set of memories sometimes called exemplars, is a central tenet of usage-based models, and thus is generally made overt in discussions of sound change from a usage-based perspective, as in this quote from Hay et al.: “The distribution of remembered pronunciations affects subsequent productions of the word” (2015:84). In these models, again, the speaker’s target is constructed from material that was experienced in the input. However, usage-based models offer more detailed accounts of two sources of disjunction between perception and production: first, exclusion of some experienced tokens from the exemplar cloud entirely (not unlike the error-

based exclusions that Martinet and Labov allow for), and second, the weighting of factors such as frequency, input recency, speaker identity, or social context in the computation of a production target. As an example of the former, Garrett & Johnson adopt a dual-representation (that is, separate perception and production representational spaces) exemplar model so that “not all instances of heard speech contribute to the pool of exemplars used in computing a memory plan” (2013:43). Questions about how production targets are derived go beyond the domain of perceptual learning. But questions about which exemplars are stored in memory in the first place, and how strongly and specifically they are stored, are, we would suggest, questions about perceptual learning.

Throughout this brief tour of some prominent discussions and models of sound change, we have seen that otherwise-distinct views share a component by which at least some of what the listener hears gets remembered over time so that it can influence that individual’s future behavior, potentially in both perception and production. These different models of sound change, then, will ultimately depend to some degree on empirical facts about perceptual learning. The interesting respects in which sound change models differ from each other often have to do with questions about exactly which aspects of the input are retained and have long-term influence, and under what circumstances. To understand exactly *how* and *what* listeners learn from perceptual input, it would be fruitful to direct our attention toward the experimental perceptual learning literature. The following subsection spells out in more detail what we might hope to learn in doing so.

2.2 Why the details of perceptual learning might matter for sound change models

In this section we sketch a slightly more detailed view of the learning processes involved in sound change. When novel phonetic input influences later behavior, that influence may take the form of adjustments in perception or adjustments in production. We begin with the question of what is involved in adjustments to production behavior because previous work, especially on change by accommodation, has already delineated finer-grained questions that offer some instructive parallels for our interest in perception. Whether we are concerned with perceptual or productive changes, we can pose questions about how long phonetic adjustments last, and how broadly phonetic adjustments are generalized.

Very broadly speaking, we take it that if momentary phonetic adjustments are to have any relationship to sound change, they eventually need

to generalize across contexts and endure over time. Auer & Hinskens cite generalization across interlocutors as the factor distinguishing the first and second steps of change-by-accommodation: “Short-term accommodation becomes long-term accommodation as soon as it permanently affects the accommodating speakers. This is the case when they transfer the innovation from direct interaction with the innovating speakers to situations in which those ‘model speakers’ fail to be the addressees” (2005:335). Reports of second dialect acquisition may be seen as exemplifying this kind of generalization in long-term accommodation (Chambers 1992; Siegel 2010; Nycz 2013, 2015). Longitudinal studies of interacting speaker pairs have similarly detected convergence beyond the interactional context (Pardo et al. 2012; Sonderegger et al. 2017).

However, the imitation literature has also made it clear that not all input is equally likely to give rise to accommodation in either the short or long term. Not only are some speakers imitated more than others (Goldinger 1998; Namy et al. 2002; Pardo et al. 2017, 2018), but also the likelihood that a particular speaker will be imitated appears to be mediated by social factors (Babel 2010; Kim et al. 2011; Babel 2012; Yu et al. 2013; Babel et al. 2014; Aguilar et al. 2016; Lewandowski & Nygaard 2018). These questions—about both who gets imitated in the first place, and whether that imitation is eventually generalized to contexts with new interlocutors—have repeatedly and explicitly been linked to sound change (recent examples include Baker et al. 2011; Yu 2013; Garrett & Johnson 2013; Babel et al. 2014; Stevens & Harrington 2014; Sonderegger et al. 2017). The experimental literature on imitation has also dedicated some attention to the linguistic dimensions of generalization in imitation, such as whether imitation generalizes to new words (Goldinger 1998; Goldinger & Azuma 2004; Nielsen 2011). Again, empirical evidence from this line of work has been discussed as directly relevant to models of sound change (Pierrehumbert 2002; Garrett & Johnson 2013; Stevens & Harrington 2014).

These detailed empirical questions about the short-term and long-term processes involved in imitation have clearly made an impact on the sound change literature. The phenomenon of imitation itself offers a basic mechanism for generating, maintaining, and spreading phonetic innovations, and then the temporal, social, and linguistic details of how imitation works have offered up ways of *constraining* that propagation in models of change. A parallel set of questions is available on the perceptual side. If a listener hears some unusual phonetic input from a speaker, will they come to expect similar input in the future from the same speaker in other linguistic or social contexts? Will they generalize that new per-

ceptual expectation to some other speakers, or even perhaps eventually shift their default perceptual norm across the board? And what are the linguistic constraints on how perceptual learning persists over time and generalizes across contexts? These questions have been at least partly addressed in the experimental literature on perceptual learning, as we will discuss at greater length in the following section. While such work often nods to possible ramifications for language change, the sound change literature itself has not fully taken up the possibilities that might be offered by a deeper understanding of how perceptual learning works.

Finally, having discussed the range of empirical questions we might ask about production changes and perception changes in response to phonetic input, we might turn to the further question of whether these interact. Are adjustments in production (in response to phonetic input) parasitic on adjustments in perception? Schertz & Clare, echoing a widely-held view, state that “in any model of sound change relying on listeners’ misattribution of cues, a change in perception would be expected to precede a change in production, both on the community and individual levels” (2019:6). We agree that *some* perceptual adjustment needs to precede any production adjustment that reflects the input—indeed, we went to some lengths to describe this relationship in Section 2.1, and this point can be seen as a driving factor in our interest in perceptual flexibility. At the same time, we note that a production change does not necessarily require full perceptual generalization. For example, in principle, a speaker could minimally adjust a speaker-specific perception norm while also adopting the new production target themselves to match that one particular speaker’s behavior, without going on to expect to hear the novel feature from other speakers. However, we do expect that people will typically have much broader perception norms than production targets, giving us some reason to guess that a broadening of a listener’s perceptual expectations is likely to often precede any change in their own behavior. We should also keep in mind the phenomenon of near-merger, where speakers in communities undergoing merger may appear to produce conservative distinctions that they do not recognize in perception (Herold 1990; Labov 1994). The perceptual questions, then, partly precede but also partly run alongside the questions about production in sound change.

3 Talker specificity in perceptual learning

One of the questions we raised in Section 2.2 is whether, or when, a listener will generalize adjusted expectations to new speakers they encounter. This question has received quite a bit of attention in the percep-

tual learning literature, where it often goes under the heading of “talker specificity.” Kraljic and Samuel put the question this way: “Do listeners learn: This odd sound is an /s/, or do they learn: This odd sound is an /s/ for Speaker X?” (2005:144). The experimental evidence on this question to date is mixed, with the answer appearing to depend on factors such as the type of contrast being tested, the acoustic similarity of the talker voices, talker gender, and talker identity. In our review of these findings, we give special attention to results suggesting that perceptual learning is talker-specific.

Talker specificity in perceptual learning is sometimes presented as eminently functional. Cutler, for example, points out that a useful perceptual adjustment in response to one speaker’s input could become an impediment to understanding the listener’s next conversational partner. For this reason, she suggests that “we should be able to adjust our phonetic categories for interpretation of the new speaker’s speech without any consequent effect on interpretation of the speech of others” (2012:397). On the other hand, generalizing an adjustment to a new speaker could be a functionally preferable choice if the listener is surrounded by speakers who share the phonetic characteristics the listener is adapting to. We mentioned second-dialect acquisition above as an example of long-term accommodation; relocation to a new dialect region is a context in which cross-talker generalization of perceptual learning would be quite handy. Indeed, perceptual learning at the level of adjusting to a entirely different accent or cluster of features is one area where generalization across talkers *has* reliably been found (Clopper & Pisoni 2004; Weatherholtz 2015; Bradlow & Bent 2008). However, when it comes to the question of how an isolated sound change takes hold in a community, single-feature perceptual learning may be the more relevant parallel. It is exactly in this literature that claims of talker specificity are found.

One of the first studies taken to provide evidence for talker specificity in perceptual learning is Eisner & McQueen (2005). They find that perceptual learning of a fricative boundary does not arise when listeners are trained on stimuli from a female voice but tested on stimuli from a male voice. They conclude that “the perceptual adjustment investigated here does not generalize across talkers” (Eisner & McQueen 2005:236). Another result that has been cited as evidence for talker specificity comes from Kraljic & Samuel (2005). This study adds an “unlearning” phase in between the training and test phases, in which listeners sometimes hear additional spoken input that either contains no cases of the critical phonemes or contains natural (non-ambiguous) instances of the critical phonemes as a form of “corrected” input. A central result of this paper is that

perceptual learning can be attenuated only with corrected unlearning input in the same talker's voice, not a different talker's voice (again, the different talker is also of the opposite gender). However, there is also some evidence in the paper suggesting that if a target pair of phonemes happens to be sufficiently acoustically similar when produced by two different voices, listeners *may* generalize across those voices to some degree.

There is also some more recent evidence weighing against previous findings of talker specificity in fricative boundary learning. Reinisch & Holt (2014) show that a lexically-guided /s/-/f/ boundary shift generalizes from a female training voice to a novel female voice, but does not generalize to a novel male voice without manipulating the acoustic similarity of the critical phonemes between the two speakers. However, this paper differs from the rest of the studies we have discussed in that the ambiguous fricatives are embedded in Dutch-accented English, so questions about the perception of foreign-accented speech also come into play. Citing the accent-adaptation studies mentioned above, they hypothesize that the presence of a foreign accent promotes generalization across talkers.

Interestingly, results on talker specificity appear to differ according to the type of phonological contrast being tested. There is evidence that perceptual learning of stop consonant boundaries is more prone to generalize across talkers than that of fricative boundaries. Kraljic & Samuel (2006) find generalization of perceptual learning across male and female talkers on a /t/-/d/ continuum. The perceptual shift in the voicing distinction also transfers to a /p/-/b/ continuum. They develop this point further in Kraljic & Samuel (2007), where they suggest that listeners learn talker-specific representations for a fricative contrast (/s/-/ʃ/) but do not do the same for a stop contrast (/t/-/d/). van der Zande et al. (2014) additionally find partial cross-talker generalization of stop contrast perceptual learning (in this case, visually rather than lexically cued).

In discussing these results, Kraljic & Samuel suggest that “when the to-be-learned phoneme highlights a temporal-voicing contrast that does not provide local, acoustic cues to speaker, as in our stop manipulations, learning will be speaker-independent. But when it highlights a spectral-place contrast that does acoustically distinguish one speaker from another, as in one of our fricative manipulations, learning is speaker-specific” (2007:3). A related proposal is found more recently in the ideal adapter model (Kleinschmidt & Jaeger 2015, 2016; Kleinschmidt 2019), which posits that listeners should generalize across speakers according to the social groupings that condition variability in speech production. Since men and women on average produce fricatives (especially /s/) with dif-

ferent spectral peaks (Jongman et al. 2000), listeners should use information about speaker gender to categorize fricatives when they encounter new talkers, and therefore should not transfer what they learn about a male talker's fricatives to a female talker or vice versa. However, if women tend to produce broadly similar fricatives, listeners might not maintain a separate mental model for the fricatives of each individual woman they encounter.

Findings of talker specificity in the fricative perceptual learning literature, as we have seen, almost always come from studies that test for generalization across male and female talkers rather than between talkers of the same gender. The fact that Reinisch & Holt (2014) find that fricative boundary learning does not generalize across genders but *does* generalize across two different female talkers, taken in conjunction with the cross-gender talker specificity results from earlier studies, suggests that generalization in these cases may actually be inhibited by voice gender mismatches, rather than tracking specific talker identities. However, the use of foreign-accented model talkers in Reinisch & Holt (2014) leaves open the possibility that the generalization across the female talkers only arose because it was supported by accent adaptation more broadly. So, while it seems that perceptual learning of fricative boundaries may not generalize across talkers of different genders, we lack a straightforward answer to the question of whether perceptual learning of fricative boundaries generalizes across talkers of the same gender.

Although this question that we've arrived at may sound somewhat narrow, we think the answer should be of real interest to historical phonologists. The evidence from stop contrasts suggests that perceptual learning *can* generalize across talkers. This is useful information for theories of sound change, because it suggests a process by which short-term, context-specific perceptual adjustments might become integrated into listeners' general expectations as the adoption of an innovation becomes widespread in a community. On the other hand, Kraljic & Samuel conclude that "when the critical sound varies along a spectral-place dimension...the system appears able to maintain multiple representations simultaneously, each for the appropriate speaker" (2007:12). While we have been discussing stops and fricatives so far, there are many other classes of sounds, notably including vowels, that reflect talker identity information. If it turned out that listeners *necessarily* build and maintain speaker-specific representations for such sounds, that might constrain our models of the relationship between perception and production in sound changes involving these phonological classes. For example, obligatory talker-specificity might be difficult to reconcile with computation of a production target dir-

ectly from a single undifferentiated pool of input tokens, especially since the speaker presumably recognizes themselves as distinct from other interlocutors.

Given the evidence amassed so far, it seems likely that generalization of perceptual learning is itself flexible, occurring across some speakers under some circumstances. The factors determining when perceptual generalization occurs, whether they are phonetic, social, or something else entirely, thus offers another dimension along which theories of change can take shape. For example, if it turns out that listeners generalize certain types of changes across same-gender but not different-gender talkers, we might look to that fact as one potential driver in the robust gender differentiation observed in many community-based studies of sound changes in progress (see e.g. Labov 2001). To probe these questions, the following section reports data from an experiment asking whether lexically-guided perceptual learning of a single fricative boundary generalizes across talkers of the same and different genders.

4 An exploration of perceptual learning with multiple talkers

In this section we present data from an exploratory perceptual learning experiment using stimuli from multiple different model talkers. We test listeners on a single female voice after training on one of four other female voices, affording four different talker-pairing opportunities to observe cross-talkers generalization. We then test different listeners on a male voice after training on one of the same four female voices, in order to compare our within-gender results to the kind of cross-gender results that claims of talker specificity have been based on.

4.1 Experimental Methods

4.1.1 Participants

Overall, 327 unique participants were recruited using the online recruitment platform Prolific (<https://app.prolific.ac/>) and were compensated for their time. Participants were restricted to be either from the United States or Canada and to have English as their first language. Of these 327, 88 (gender information not collected) participated in the continuum pilot described below, 121 (57 women, 63 men, and one person of another gender identity) participated in a perceptual learning experiment testing categorization of fricatives from a female voice, and 118 (44 women, 73 men, and one person of another gender identity) participated in a perceptual learning experiment testing categorization of fricatives

from a male voice. Two participants were excluded for reporting that they were from a location outside the US and Canada.

4.1.2 Materials

Five female speakers and one male speaker, all native speakers of American English, recorded stimuli for the experiments presented here. They will be referred to throughout as (female speakers) F1, F2, F3, F4, and F5, and (male speaker) M1. Of these, F1 and M1 serve as test speakers and F2-F5 serve as training speakers. For each speaker, we first created an [s]-[f] continuum. To create the continua, we selected tokens of [ɛs] and [ɛf] recorded by each speaker, then used a custom Praat (Boersma & Weenink 2002) script to blend the [s] and [f] together in five-percentage-point increments from (20% [f], 80% [s]) to (95% [f], 5% [s]). For speakers F2-F5, we then used a preliminary rating task to select the optimally ambiguous fricative from within the continuum. In the preliminary rating task, we randomly presented each increment five times to 14-15 participants per continuum and asked them to indicate whether each sound they heard was an 'S' or an 'F' using their keyboard. The maximally ambiguous token across participants was then selected by determining which blended token was nearest to the 50% perceptual boundary between /f/ and /s/. The chosen maximally ambiguous points for speakers F2-F5 respectively had [f] proportions of 55%, 65%, 65%, and 80%.

The spectral centers of gravity (COG) averaged across the full fricative duration for the selected maximally ambiguous fricative from each training voice, as measured in Praat, is given in Table 1. Table 1 also includes the same measure for the test speaker fricatives that turned out to be maximally ambiguous in the experiment pre-test (as described below in Section 4.1.3), which happened to be 75% for both test speakers. Note that the female test voice (F1) has a much lower COG than any of the female training voices (F2-F5), while the male test voice (M1) happens to fall within the range of the female training voices.

Table 1: Average spectral center of gravity (Hz) for the maximally-ambiguous fricative for each training and test voice

| Test | | Training | | | |
|------|------|----------|------|------|------|
| M1 | F1 | F2 | F3 | F4 | F5 |
| 7839 | 5450 | 6902 | 8373 | 8713 | 8617 |

The maximally ambiguous token for each training speaker was spliced into twenty words ending in /f/ from the same talker by first removing the final /f/ and then inserting the ambiguous token. These 20 words with ambiguous fricatives (/ʔ/-final words) were used along with 20 unmanipulated /s/-final words, 10 fillers, and 50 non-words to create the lexical decision task. Stimuli were varied by the number of syllables, such that for the both /ʔ/- and /s/-final words, there were 10 monosyllabic, 5 disyllabic, 4 trisyllabic, and 1 tetrasyllabic words. Filler and non-word stimuli were also varied as to the number of syllables in roughly the same distribution. The fricative /f/ never occurred in any position except as the ambiguous /ʔ/ in the 20 /ʔ/-final words. Likewise, the fricative /s/ only appeared in the 20 /s/-final words. The full list of training words and non-words can be found in the Associated Materials.

4.1.3 Procedures

The experimental platform Ibx was used, along with the PennController system (Zehr & Schwarz 2018), to implement the experimental presentation. The experiment with F1 as the test voice was run separately from the experiment with M1 as the test voice. Within those two versions of the experiment, participants were randomly assigned to one of four experimental conditions corresponding to the four training voices, F2-F5. Each participant completed three phases in sequence: a pre-test categorization task with the assigned test voice (either F1 or M1), followed by a training lexical decision phase with the assigned training voice (one of F2-F5), followed by a post-test categorization task with the same test voice as the pre-test. Perceptual learning within this experiment is thus assessed on a within-subject basis, in comparison to the more typical between-subjects design in previous work.

After consenting to participate, participants were told they would be hearing sounds that they should classify as either an /s/ or /f/ using the respective keyboard keys. They were told to respond as quickly and accurately as possible and were additionally informed that there would be two additional phases after the pre-test, but were not informed of the content of the following phases. After two practice trials, the categorization pre-test contained a total of 120 trials in a set random order with a random ISI between 400-600ms between trials. The trials consisted of 10 presentations each of 12 steps of the [s]-[f] continuum from the assigned test speaker (either F1 or M1), starting at (40% /f/, 60% /s/) and ending at (95% /f/, 5% /s/). We start the continuum at a fairly high /f/ propor-

tion because the noisy sibilant /s/ dominates the perception of less /f/-ful steps.

Following the categorization pre-test, participants were presented with instructions for a continuous auditory lexical decision task. They were again told that they would hear a set of sounds and should respond as quickly and accurately whether the sound they heard was a word (using the 'm' key) or not a word (using the 'z' key). After two practice trials, the continuous auditory lexical decision task contained 100 stimuli presented in a set random order with a random ISI of 400-600 ms between trials. Of these stimuli, 20 were /?/-final words, 20 were /s/-final words, 10 were additional filler real words, and the remaining 50 were nonwords (as described in Section 4.1.2).

The final phase of the experiment, the categorization post-test, was a direct copy of the categorization pre-test, with instructions referencing the pre-test's instructions. After completion of the post-test, a small demographic questionnaire was administered, allowing participants to provide their self-identified gender, age, country of origin, and comments concerning the experiment.

Additional participant exclusions were based on experimental performance. 19 participants in the F1 test experiment and 14 participants in the M1 test experiment were excluded either for having an accuracy rate of lower than 70% on the lexical decision task or for having fewer than 50 percentage points spread in syllable classification rates between continuum endpoints. There were 25-26 participants in every condition analyzed after these exclusions.

4.2 Results

Because the goal of the perceptual learning paradigm is to induce listeners to shift their perceptual boundary between phonemes, a natural way to present the results is by plotting the categorization boundary on the /s/-/f/ test continuum between pre-test and post-test. Figure 1 shows the pre-test (red) and post-test (green) rates at which participants gave /f/ responses on the /s/-/f/ continuum for M1, the male test voice. Each facet represents data from a different training voice condition - recall that all four training speakers were female. Impressionistically, what we see in Figure 1 is minimal change in listener behavior between pre-test and post-test. This result is consistent with the evidence for talker specificity surveyed in Section 3; recall that all such evidence came from experiments which also used training and test voices from speakers of different genders.

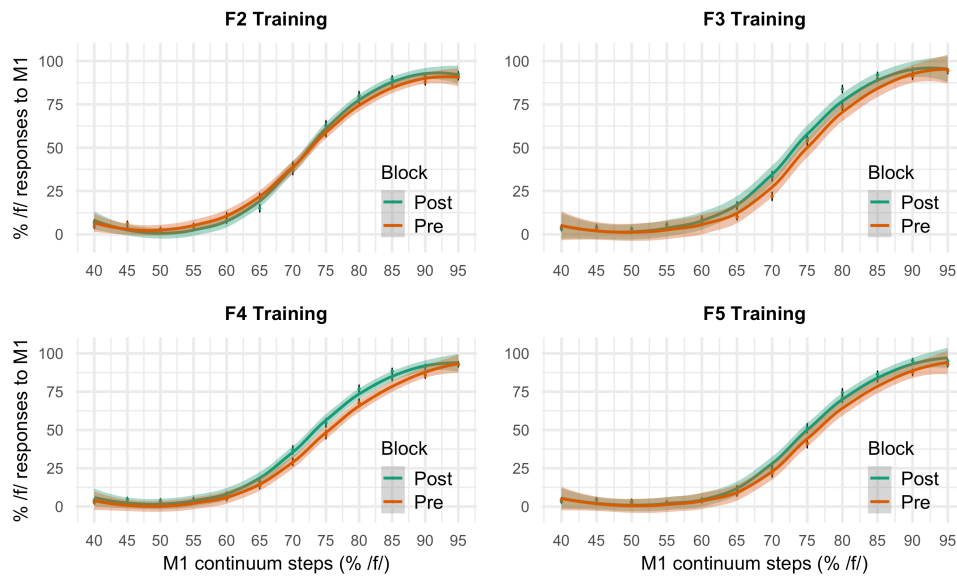


Figure 1: Pre-test and post-test classification for M1 continuum (F2-F5 training)

Figure 2 shows the parallel analysis of the data from participants responding to the F1 test voice. In this case, it appears there is a much more substantial change from pre-test to post-test, with listeners becoming more likely to classify ambiguous steps in speaker F1's continuum as /f/ after being exposed to /f/-biased stimuli from a *different* female speaker. This result is not consistent with the view that perceptual learning is talker specific.

At this point we might normally proceed to statistical modeling and a tidy conclusion, but further exploratory analysis revealed a wrinkle in these results that take us on a detour away from that path. An analysis of the time course of the categorization data reveals that across all conditions in both experiments, participants begin shifting their perceptual boundary in the expected direction during the categorization pre-test—*before* being exposed to any intended training stimuli. Figure 3 shows what this looks like in the pre-test categorization data for a single condition. The first 10% of data is seen in the darkest line, the next 10% in the next darkest line, and so on. As more pre-test classification data is presented to the participants, the likelihood of an /f/ response rate increases. We see this pattern across all of the pre-test categorization phases, which raises questions about the interpretability of the shifts that we see from pre-test to post-test.

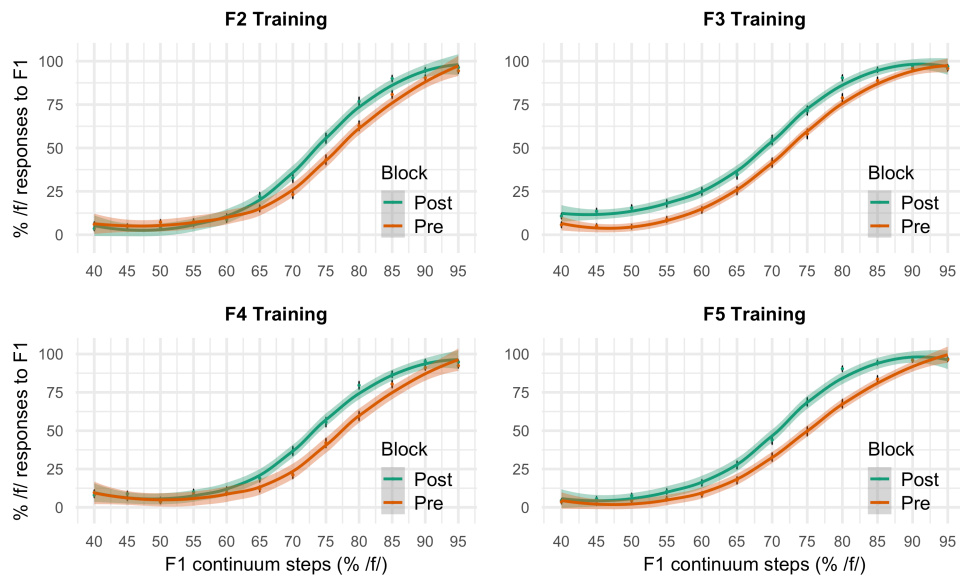


Figure 2: Pre-test and post-test classification for F1 continuum (F2-F5 training)

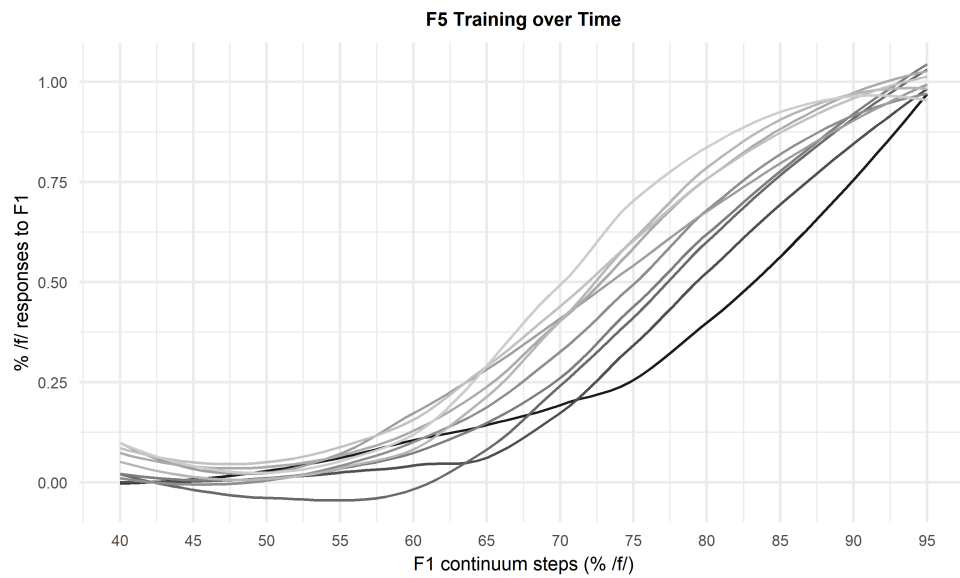


Figure 3: Pre-test classification over time (in F5 training condition); lines become lighter over time

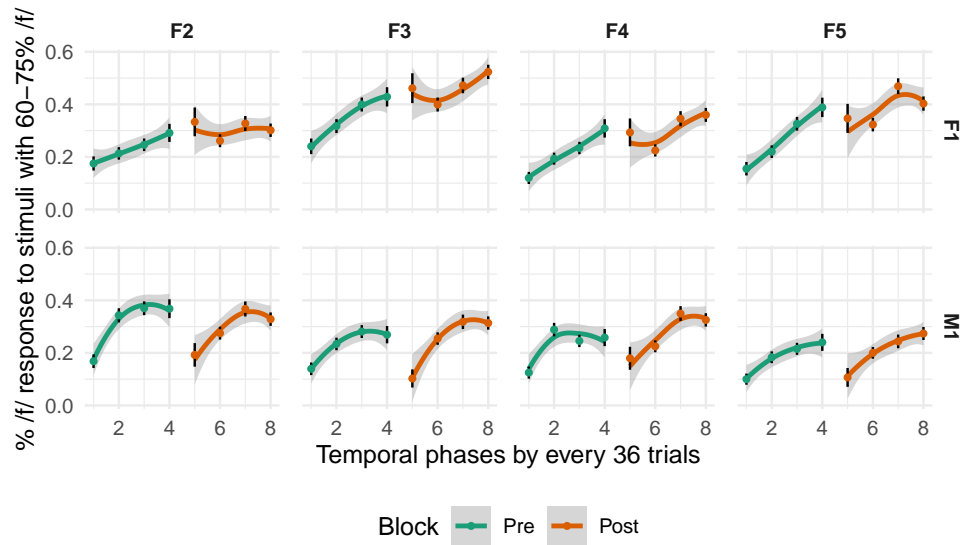


Figure 4: Pre-test and post-test syllable classification by trial block

Figure 4 breaks the pre-test and post-test trial sequences into bins of 36 trials each and presents the average /f/ response rate for the continuum steps between 60-75% [f]. The basic patterns that we wish to highlight in this graph are as follows: 1) that listeners in all conditions give gradually more /f/ responses over the course of the pre-test, 2) that listeners responding to the female test voice begin their post-test at an /f/ response rate comparable to their /f/ response rate at the end of the pre-test, 3) that in contrast, listeners responding to the male test voice, after the intervening female-voice training phase, go back to an /f/ response rate comparable to the start of the pre-test, and 4) that listeners in all conditions continue to increase their /f/ response rate over the course of the post-test as well.

We attribute the pre-test shift behavior to the fact that the test continua from both M1 and F1 are asymmetrical: listeners hear more unambiguous /s/ steps than unambiguous /f/ steps. It appears that the listeners are demonstrating a more basic behavior that might be understood as reflecting a *range effect* (Brady & Darwin 1978; Rosen 1979; Keating et al. 1981), involving interpretation of the continuum endpoints as phonemic anchors, or a *frequency effect*, involving something like a bias toward hearing each option an equal number of times in a two-alternative

forced choice task.¹ We believe it is still correct to view this as a form of perceptual learning, albeit not lexically-guided, because it still involves a change in speech processing subsequent to exposure to acoustic stimuli. Whatever its motivation, though, its appearance here does complicate the interpretation of the results in Figures 1 and 2. The direction of the pre-training shift is the same as the expected direction of the lexically-guided perceptual learning, so it becomes difficult to isolate our intended effect.

More precisely, this ambiguity allows for two possible interpretations of why the learning behavior seen in Figure 4 is differentiated by test voice gender. One interpretation of Figure 4 is that the switches between male and female voices are triggering a reset back to a default expectation, whereas the switches between distinct female voices are not. The primary driver of the boundary shift over both pre-test and post-test in Figure 4, then, would be the learning response to the continuum asymmetry, with a different-gender voice interrupting and resetting that process to start from scratch in the post-test. However, it is also still possible that the intended lexically-cued manipulation is triggering additional learning. So another possible interpretation is that the resetting takes place in both M1 and F1 test contexts, the asymmetry-response continues in the post-test in both contexts, and the additional boost to /f/-response rates in the F1 post-test is due to the generalization of lexically-guided perceptual learning across the female voices but not from female to male.

Despite this entanglement of slightly different learning processes, the results as we see them in Figure 4 still bear directly on the larger issues we raised in Sections 2 and 3. Regardless of which of our interpretations is correct, we see that listener behavior is substantially different for two same-gender talkers as it is for two different-gender talkers. This reinforces that the lack of same-gender cross-talker studies is a gap in the literature, which is of potential importance for theories of sound change for the reasons we discussed in Section 3. In the following section, we discuss how our general pattern of results relates to models of perceptual learning.

5 Discussion

As we laid out in Section 1, our three goals in this paper were to articulate the connection between mechanisms of perceptual flexibility and models of sound change, to review the evidence for talker specificity in percep-

¹ Thanks to Joe Toscano for pointing us to the literature on range and frequency effects.

tual learning, and to report novel experimental data that pertains to the talker specificity question. With respect to the first of these goals, we argued in Section 2 that the integration of perceptual input into longer-term linguistic expectations and production behavior is a necessary component for a range of different approaches to sound change. As a field, we do not have a full picture of the learning and generalization processes that underlie the adjustment of both perception norms and production targets in sound change. We moved on to our second goal in identifying questions about talker specificity in perceptual learning as one area where empirical answers about such processes could prove useful in understanding sound change. Section 3 reviewed what is and isn't known on this point. We pointed out that it is unknown whether fricative boundary retuning—and perhaps by extension, retuning for other spectrally-cued sounds such as vowels—does or does not generalize across different speakers of the same gender. We presented the experiment in Section 4 with the intention of answering that question.

In the experimental data, we observed large perceptual shifts from pre-test to post-test in all four of the same-gender voice pair conditions, and minimal shifts from pre-test to post-test in all four of the different-gender voice pair conditions. Whether we interpret this pattern as arising from the lexically-guided training phase or from learning of more basic statistical properties of the test continuum, we can still consider whether talker-specific representations might be at play. One possibility we cannot rule out is that listeners might have mistaken the different female voices as coming from a single speaker. This possibility is, of course, why different-gender pairs are typically used in this paradigm in the first place. A more definitive test might involve the inclusion of disambiguating information about speaker identity, as in van der Zande et al. (2014); since they found that showing the model talkers' faces did *not* block cross-talker generalization of stop continuum learning, we are hesitant to chalk up the gender differences in our results entirely to voice confusion. The fricative-specific phonetic differences between F1 and the other female voices also weigh against this interpretation.

If we set aside the voice confusion possibility, the patterns of generalization and specificity in our experiment are not straightforwardly compatible with a view that listeners are building and maintaining talker-specific acoustic representations from the fricative input they hear. Even on the interpretation where only the statistical properties of the test continuum are being learned, the shift seen in the M1 pre-test is discarded after hearing a female voice. One alternative to talker-specificity that has been put forward in the literature is that generalization is supported by

acoustic similarity across voices. Because two voices of the same gender are likely to be more similar, on average, generalization rooted in acoustic overlap might also be expected to give rise to the appearance of within-gender but not across-gender generalization. However, recall that in our experimental materials, the male test voice is more similar to the female training voices than the female test voice is (at least in terms of the ambiguous fricative COGs). This suggests that the generalization patterns are not being driven by the local (within-experiment) informativity of the target phonetic property. The ideal adapter model might prove more useful for understanding the gender differences in our data because it gives weight to a listener's accumulated experiences with the informativity of social groupings (such as gender) in speech perception. As a result, it allows listener knowledge of *typical* gendered phonetic patterns to exert an influence on perceptual generalization behavior even when the listener hears voice combinations that don't instantiate the same patterns.

Theoretical psycholinguistic accounts of perceptual learning continue to advance, and in doing so create an ongoing need for experimental work to evaluate increasingly fine-grained predictions about their mechanistic details. We believe historical phonologists might find it fruitful to keep abreast of how experimental work in this vein is developing, or even to pursue collaborative experimental work in order to draw focus to aspects of perceptual learning that are most relevant to questions about sound change. Experimental results from a fuller set of phoneme contrast types, for example, might be linked to typological patterns of change in the historical record. Experiments on how the distributional properties of socioindexical phonetic information shape perceptual learning might connect with the well-developed literature on how sound changes spread through socially-stratified speech communities. The questions we raised at the end of Section 2.2, about the extent to which changes in speakers' production behavior are dependent on perceptual learning being generalized to different degrees, represent another area where a great deal of additional experimental inquiry is needed and a great deal might be learned from such work. Theories of sound change have long benefited from experimental insights on topics such as compensation for coarticulation and the many factors constraining phonetic imitation. We are optimistic that

the domain of perceptual learning offers similar possibilities that have not yet been uncovered.

Comments invited

PiHPh relies on post-publication review of the papers that it publishes. If you have any comments on this piece, please add them to its comments site. You are encouraged to consult this site after reading the paper, as there may be comments from other readers there, and replies from the author. This paper's site is here:

<https://doi.org/10.2218/pihph.5.2020.4439>

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant Number 1627972, "Cognitive characteristics of the leaders of language change." Thanks to Elisha Cooper for her contributions in experimental design and implementation and to our many wonderful undergraduate research assistants. Thanks also to Sabriya Fisher, Ava Irani, and Betsy Sneller for lending their voices. The ideas in this paper have been improved by feedback from audiences at CUNY Graduate Center, Ludwig Maximilian University of Munich, and especially the Edinburgh Symposium on Historical Phonology. We are also grateful to Dave Embick and Joe Toscano for their input.

Associated material

Lexical Decision Stimuli

| Ambiguous F-Final Words | Non-Ambiguous S-Final Words | Fillers | Non-Words (ARPabet) | |
|------------------------------------|--|----------------|----------------------------|------------------------------|
| autograph | ambiguous | create | B AH1 L CH | K AH1 G D AH0 T |
| behalf | ambitious | grind | B IH1 NG K | K EH1 T B IH0 M |
| belief | arthritis | helicopter | B IH1 P | K IH1 G B EH0 T** |
| brief | bliss | implode | B L AY1 | K IY1 T |
| chef | bypass | laundry | B R AE1 M | K L AH1 M |
| chief | cactus | moat | CH AO1 R N | K L AO1 M |
| choreograph | chess | opinion | D AE1 CH | L AA1 M N AA0 P |
| cliff | class | remember | D IH1 B R AH0 D | M AA1 T |
| cough | contagious | tab | D IH1 T R OW0 P EH2 N | M AA1 T Y AA0 B |
| dandruff | courageous | town | D R AE1 D | M AE1 M P |
| deaf | harass | | D R IH1 JH | N EH1 M |
| giraffe | mass | | G AE1 N D | N IY1 K |
| golf | niece | | G AE1 T K IH0 P | N UH2 G L AH0 N JH EH1 K IY0 |
| handkerchief | pace | | G R AA1 K | P EY1 P |
| pocketknife | piece | | G UH1 P L IH0 M P AA2 B | P L EH1 L |
| rough | plus | | HH AE2 P K AE1 G D IY0 | P L IH1 N |
| safe | price | | HH OW1 K | P R AE1 N |
| tariff | tennis | | JH AY1 T | R AO1 R B |
| waterproof | toss | | JH EH1 R P | R IH1 T JH OW0 M |
| wife | walrus | | JH OW1 T* | R OW1 K |
| | | | SH AA1 G | T R IH1 N |
| | | | T AA1 M N AH0 D | W AE1 P M IH0 G |
| | | | T AA1 R K | W AH1 M |
| | | | T EH1 B | W AH1 NG |
| | | | T R EH1 NG K | Y AE1 N D |

*replaced with 'lahlb' for one talker

**accidentally read as 'sihgbeht' for one talker

Author contact details

Meredith Tamminga
University of Pennsylvania
3401-C Walnut St, Suite 300C
Philadelphia, PA 19104
USA
tamminga@ling.upenn.edu

Robert Wilder
University of Pennsylvania
3401-C Walnut St, Suite 300C
Philadelphia, PA 19104
USA
rwilder@sas.upenn.edu

Wei Lai
University of Pennsylvania
3401-C Walnut St, Suite 300C
Philadelphia, PA 19104
USA
weilai@sas.upenn.edu

Lacey Wade
University of Pennsylvania
3401-C Walnut St, Suite 300C
Philadelphia, PA 19104
USA
laceya@sas.upenn.edu

References

- Aguilar, Lauren, Geraldine Downey, Robert Krauss, Jennifer Pardo, Sean Lane & Niall Bolger. 2016. A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality* 84(2). 165–177. doi:10.1111/jopy.12149.
- Auer, P. & F. Hinskens. 2005. The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens & P. Kerswill (eds.), *Dialect change. the convergence and divergence of dialects in con-*

- temporary society*, 35–57. Cambridge University Press. doi:10.1017/cbo9780511486623.015.
- Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39(4). 437–456. doi:10.1017/s0047404510000400.
- Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40. 177–189. doi:10.1016/j.wocn.2011.09.001.
- Babel, Molly, Grant McQuire & Jamie Russell. 2014. Whose sound changes do we follow? Selective attention to ingroups as a mechanism for sound change. Paper presented at Sound Change in Interacting Human Systems, May 29.
- Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23. 347–374. doi:10.1017/s0954394511000135.
- Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85(4). doi:10.1353/lan.0.0165.
- Bertelson, Paul, Jean Vroomen & Béatrice de Gelder. 2003. Visual recalibration of auditory speech identification: A McGurk after-effect. *Psychological Science* 14(6). 592–597.
- Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
- Boersma, Paul & David Weenink. 2002. Praat, a system for doing phonetics by computer. *Glott International* 5.
- Bradlow, Ann R. & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106. 707–729. doi:10.1016/j.cognition.2007.04.005.
- Brady, Susan A. & C. J. Darwin. 1978. Range effects in the perception of voicing. *Journal of the Acoustical Society of America* 63(5). 1556–1558.
- Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261–290. doi:10.1017/s0954394502143018.
- Chambers, J.K. 1992. Dialect acquisition. *Language* 68(4). 673–705.
- Clarke, Constance M. & Merrill F. Garrett. 2004. Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* 116(6). 3647–3658. doi:10.1121/1.1815131.
- Clopper, Cynthia G. & David B. Pisoni. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47(3). 207–239.
- Connine, Cynthia M. & Laura M. Darnieder. 2009. Perceptual learning of co-articulation in speech. *Journal of Memory and Language* 61. 412–422.

- Cutler, A., K.M. McQueen, S. Butterfield & D. Norris. 2008. Prelexically-driven perceptual retuning of phoneme boundaries. In *Proceedings of Interspeech 2008*, Brisbane, Australia.
- Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.
- Eisner, Frank & James M McQueen. 2005. The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics* 67(2). 224–238.
- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 110.
- Garrett, Andrew & Keith Johnson. 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, 51–97. Oxford University Press.
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279.
- Goldinger, Stephen D. & Tamiko Azuma. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review* 11(4). 716–722.
- Hay, Jennifer & Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92(2). 298–330. doi:10.1353/lan.2016.0036.
- Hay, Jennifer B., Janet B. Pierrehumbert, Abby Walker & Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139. 83–91.
- Herold, Ruth. 1990. *Mechanisms of merger: the implementation and distribution of the low back merger in Eastern Pennsylvania*: University of Pennsylvania dissertation.
- Jesse, Alexandra & Elina Kaplan. 2019. Attentional resources contribute to the perceptual learning of talker idiosyncrasies in audiovisual speech. *Attention, Perception, & Psychophysics* 81. 1006–1019. doi:10.3758/s13414-018-01651-x.
- Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Maria-Josep Solé, Patrice Speeter & Manjari Ohala (eds.), *Experimental approaches to phonology: In honor of John Ohala*, Oxford.
- Jongman, Allard, Ratrete Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108(3). 1252. doi:10.1121/1.1288413.
- Keating, Patricia A., Michael J. Mikós & William F. Ganong, III. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America* 70(5).

- 1261–1271.
- Kim, Midam, William S. Horton & Ann R. Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2(1). doi:10.1515/labphon.2011.004.
- Kleinschmidt, Dave F. 2019. Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience* 34(1). 43–68.
- Kleinschmidt, Dave F. & T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122(2). 148–203. doi:10.1037/a0038695.
- Kleinschmidt, David & T. Florian Jaeger. 2016. What do you expect from an unfamiliar talker? In John Trueswell, Anna Papafragou, Daniel J. Grodner & Daniel Mirman (eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.
- Kraljic, Tanya & Arthur G. Samuel. 2005. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51(2). 141–178.
- Kraljic, Tanya & Arthur G. Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13(2). 262–268.
- Kraljic, Tanya & Arthur G. Samuel. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56(1). 1–15.
- Labov, William. 1994. *Principles of linguistic change, vol. 1. internal factors*. Oxford: Blackwell.
- Labov, William. 2001. *Principles of linguistic change, vol. 2. social factors*. Oxford: Blackwell.
- Lewandowski, Eva M. & Lynne C. Nygaard. 2018. Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America* 144(2). 620–633. doi:10.1121/1.5038567.
- van Linden, S. & J. Vroomen. 2007. Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance* 33. 1483–1494.
- Martinet, André. 1952. Function, structure, and sound change. *WORD* 8(1). 1–32.
- McAuliffe, Michael. 2015. *Attention and salience in lexically-guided perceptual learning*: University of British Columbia dissertation.
- Namy, Laura L., Lynne C. Nygaard & Denise Sauerteig. 2002. Gender differences in vocal gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21. 422–432. doi:10.1177/026192702237958.

- Niedzielski, Nancy & Howard Giles. 1996. Linguistic accommodation. In J. Holmes, Hans Goebel, P. H. Nelde, Z. Stary & W. Wölck. (eds.), *Contact linguistics: An international handbook of contemporary research*, 332–342. Walter de Gruyter.
- Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2). 132–142.
- Norris, Dennis, James M. McQueen & Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204–238.
- Nycz, Jennifer. 2013. Changing words or changing rules? second dialect acquisition and phonological representation. *Journal of Pragmatics* 52(49-62).
- Nycz, Jennifer. 2015. Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass* 9. 469–482.
- Ohala, John J. 1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (eds.), *Papers from the parasession on language and behavior*, Chicago: Chicago Linguistics Society.
- Ohala, John J. 1993. Coarticulation and phonology. *Language and Speech* 36. 155–170.
- Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes & Robert M. Krauss. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40(1). 190–197. doi:10.1016/j.wocn.2011.10.001.
- Pardo, Jennifer S., Adelya Urmanche, Sherilyn Wilman & Jaclyn Wiener. 2017. Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* 79(2). 637–659. doi:10.3758/s13414-016-1226-0.
- Pardo, Jennifer S., Adelya Urmanche, Sherilyn Wilman, Jaclyn Wiener, Nicholas Mason, Keagan Francis & Melanie Ward. 2018. A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics* 69. 1–11. doi:10.1016/j.wocn.2018.04.001.
- Paul, Hermann Otto Theodor. 1880. *Prinzipien der sprachgeschichte*. Halle: Max Niemeyer Verlag.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In *Laboratory Phonology VII*, 101–139. Berlin: Mouton de Gruyter.
- Reinisch, Eva & Lori L. Holt. 2014. Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance* 40(2). 539–555. doi:10.1037/a0034409.
- Repp, Bruno H & Alvin M Liberman. 1984. Phonetic category boundaries are flexible. *Haskins Laboratories Status Report on Speech Research (SR-77/78)* 31–53.

- Rosen, Stuart Michael. 1979. Range and frequency effects in consonant categorization. *Journal of Phonetics* 7. 393–402.
- Samuel, Arthur G & Tanya Kraljic. 2009. Perceptual learning for speech. *Attention, Perception, & Psychophysics* 71(6). 1207–1218.
- Schertz, Jessamyn & Emily J. Clare. 2019. Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science* doi:10.1002/wcs.1521.
- Shockley, Kevin, Laura Sabadini & Carol Fowler. 2004. Imitation in shadowing words. *Attention, Perception, and Psychophysics* 66(3). 422–429.
- Siegel, Jeff. 2010. *Second dialect acquisition*. Cambridge: Cambridge University Press.
- Sonderegger, Morgan, Max Bane & Peter Graff. 2017. The medium-term dynamics of accents on reality television. *Language* 93(3). 598–640.
- Stevens, Mary & Jonathan Harrington. 2014. The individual and the actuation of sound change. *Loquens* 1(1). e003. doi:10.3989/loquens.2014.003.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford and New York: Basil Blackwell.
- Vroomen, J. & M. Baart. 2009. Recalibration of phonetic categories by lipread speech: Measuring aftereffects after a 24-hour delay. *Language and Speech* 52. 341–350.
- Vroomen, J., S. van Linden, B. de Gelder & P. Bertelson. 2007. Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia* 45(572–577).
- Vroomen, J., S. van Linden, M. Keetels, B. de Gelder & P. Bertelson. 2004. Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication* 44. 55–61.
- Weatherholtz, Kodi. 2015. *Perceptual learning of systemic cross-category vowel variation*: The Ohio State University dissertation.
- Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, chap. 10, 201–227. Oxford University Press.
- Yu, Alan C. L., Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PLoS ONE* 8(9). e74746. doi: 10.1371/journal.pone.0074746.
- van der Zande, Patrick, Alexandra Jesse & Anne Cutler. 2014. Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics* 43. 38–46. doi:10.1016/j.wocn.2014.01.003.

- Zehr, Jeremy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). Retrieved from <https://doi.org/10.17605/OSF.IO/MD832>.
- Zellou, Georgia, Rebecca Scarborough & Kuniko Nielsen. 2016. Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America* 140(5). 3560–3575.