Fall 2011

Perceptual learning in speech reveals pathways of processing

Cheyenne Michele Munson
University of Iowa

PERCEPTUAL LEARNING IN SPEECH REVEALS PATHWAYS OF PROCESSING

by

Cheyenne Michele Munson

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Psychology
in the Graduate College of
The University of Iowa

December 2011

Thesis Supervisor: Associate Professor Bob McMurray

ABSTRACT

Listeners use perceptual learning to rapidly adapt to manipulated speech input. Examination of this learning process can reveal the pathways used during speech perception. By assessing generalization of perceptually learned categorization boundaries, others have used perceptual learning to help determine whether abstract units are necessary for listeners and models of speech perception. Here we extend this approach to address the inverse issue of specificity. In these experiments we have sought to discover the levels of specificity for which listeners can learn variation in phonetic contrasts. We find that (1) listeners are able to learn multiple voicing boundaries for different pairs of phonemic contrasts relying on the same feature contrast. (2) Listeners generalize voicing boundaries to untrained continua with the same onset as the trained continua, but generalization to continua with different onsets depends on previous experience with other continua sharing this different onset. (3) Listeners can learn different voicing boundaries for continua with the same CV onset, which suggests that boundaries are lexically-specific. (4) Listeners can learn different voicing boundaries for multiple talkers even when they are not given instructions about talkers and their task does not require talker identification. (5) Listeners retain talker-specific boundaries after training on a new boundary for a second talker, but generalize boundaries across talkers when they have no previous experience with a talker. These results were obtained using a new paradigm for unsupervised perceptual learning in speech. They suggest that models of speech perception must be highly flexible in order to accommodate both specificity and generalization of perceptually learned categorization boundaries.

Abstract Approved: ______________________________
Thesis Supervisor

____
Title and Department

____
Date

PERCEPTUAL LEARNING IN SPEECH REVEALS PATHWAYS OF PROCESSING

by

Cheyenne Michele Munson

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Psychology
in the Graduate College of
The University of Iowa

December 2011

Thesis Supervisor: Associate Professor Bob McMurray

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Cheyenne Michele Munson

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Psychology at the December 2011 graduation.

Thesis Committee: ___________________________
Bob McMurray, Thesis Supervisor

____
Gregg Oden

____
Prahlad Gupta

____
Karen Kirk

____
Eliot Hazeltine

____
Jill Beckman

ACKNOWLEDGEMENTS

Like any scientific production, this dissertation has been a collaborative effort and there are many people who deserve my thanks. First my advisor Bob McMurray, who has been a fantastic mentor and who knew the meaning of what I was doing before I did myself. Thanks for staying up much later than you should have needed to in order to finish all the editing.

The other members of the MacLab deserve my thanks for their help with various tasks like piloting stimuli (over and over and over again) and making tedious VOT measurements. I especially want to thank our lab manager Dan McEchron for all the scheduling he did that made it possible for me to get these experiments run.

I would also like to thank my circle of colleagues. Those at Iowa include Katy Mueller, Gwyneth Rost, Christine Zeimer, Kim Halvorson, Keith Apfelbaum, Ashley Farris-Trimble, and Marcus Galle. You were there to talk science and you were there for everything else. My new colleagues at Illinois (including Simon Fischer-Baum, Maureen Gillespie, and Jessica Love) have made me feel just as at home in a new place.

My committee members Prahlad Gupta, Eliot Hazeltine, Karen Kirk, Jill Beckman, and Gregg Oden have provided feedback critical to my work, on this dissertation and otherwise. Gregg was particularly helpful for my understanding of FLMP, and Jill brought a great number of typos to my attention. The rest of the faculty and staff in the psychology department at Iowa have been important too, helping me stick to deadlines and building an intellectual community that I have been proud to be a part of.

There are many people I don't know at all who have helped me with this dissertation, especially those who have contributed to the open-source software products that I have used (Praat, R, and Latex). I appreciate the work that has gone into these, and all the help that is available to those of us still trying to figure things out.

Then of course there are the people I know the best, my family and friends who have been supportive of this endeavor even when they didn't really know what I was working on. My parents are still hoping that I'll move closer to home someday, but I know they understand why I'm so far away, and they will always be there for me.

Finally there is Joe, who did know what I was working on, and who helped me through it all. You put up with a lot and were still there at the end. I cannot thank you enough.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

When learning a language, acquisition of all the categories of sounds it uses is a critical problem faced by all learners. Figuring out which differences between sounds are meaningful for language is typically considered to occur during infancy as infants are exposed to their language (Werker & Tees, 1984; Werker & Curtin, 2005). However, although these sound categories must be learned early, they must also remain malleable even in adults. This is critical given that pronunciations of phonemes change over time, both for individuals (Bauer, 1985; Harrington, Palethorpe, & Watson, 2000) and for entire language communities (Cox, 1999; Labov, 1994; Watson, Maclagan, & Harrington, 2000). Listeners must thus be able to accommodate these changes. Moreover, the malleability of phonetic categories may also help listeners adjust to individual variability in speech production and to accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Kraljic, Samuel, & Brennan, 2008; Magen, 1998; Sidaras, Alexander, & Nygaard, 2009), as short-term learning mechanisms allow them to determine the relevant phonetic categories while speaking to someone with an accent. This type of plasticity has been supported by substantial work over the last few years, which has shown that in laboratory settings, adult listeners use a process of perceptual learning to quickly adjust their phoneme categories to better match the input they hear (e.g., Norris, McQueen, & Cutler, 2003; McQueen & Mitterer, 2005; McQueen, Cutler, & Norris, 2006; Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Kraljic et al., 2008; Clarke & Luce, 2005; Clarke-Davidson, Luce, & Sawusch, 2008; Sjerps & McQueen, 2010).

A key question regarding perceptual learning in speech perception is the degree to which it generalizes. For example, after learning how a single talker produces a given category, do listeners generalize this to all talkers, assuming that everyone produces that sound in the same way? Similarly, if they learn how voicing is instantiated at one place of articulation (e.g., b/p) do they generalize it to others (k/g)? Generalization across talkers in particular may be helpful for adapting to groups of similar talkers, but there must also be some degree of specificity to perceptual learning; otherwise listeners would be unable to cope with individual variability or adapt to multiple accents. A number of studies have addressed generalization of perceptual learning. In these studies, the question is whether listeners generalize what they learn to new words, talkers, and phonemes. The inverse question of specificity, on the other hand, has received little attention. It is largely unknown whether it is possible for listeners to learn shifts that apply only to specific talkers, phonemes, or words. Indeed, if such specificity can be found, it would allow us to pinpoint where in the language-processing stream this learning occurs. Moreover, examining the specificity of perceptual learning along with generalization may provide us with additional information about the degree of abstraction present in the speech perception system. Thus, the goal of this dissertation is to determine what levels of processing are affected by perceptual learning, to discover the levels of specificity for which listeners can learn variation in phonetic contrasts, and to examine the implications of these results for various models of speech perception.

In the remainder of this chapter I will first discuss the different levels at which the perceptual system might be sensitive to perceptual learning, the different patterns generalization and specificity in learning that may occur, and which models of speech perception would be consistent with different patterns of results. Next I will review some basic findings on perceptual learning and discuss the existing literature on the generalization and specificity of perceptual learning for words, phonemes, and talkers. The review ends with the specific aims of the experiments included in this dissertation. Finally, the following chapters present a methodological overview and a series of six experiments examining the level at which specific or generalized perceptual learning can be observed in speech perception.

1.1 Theoretical Implications of Perceptual Learning

One of the reasons that perceptual learning is worth examining in detail is that it may provide key insights about what levels of processing comprise the speech perception system, and the degree of abstraction necessary at these different levels. While models of speech perception were, for the most part, not developed to address questions about learning, experiments on learning may still tell us something about what type of architecture a model of speech perception needs. This approach assumes that learning could, at some point, be incorporated into any of the models. While this has not been demonstrated for many models of speech perception, at least some models that were not built for learning have had learning incorporated quite successfully. The best example of this is the extension of the TRACE model (McClelland & Elman, 1986) developed by Mirman, McClelland, and Holt (2006). Interactive activation and Hebbian learning are used to adjust pre-lexical representations based on feedback from the lexicon, which allows the model to explain a number of perceptual learning effects, including lexically driven perceptual learning for shifted phoneme category boundaries and generalization of perceptual learning across talkers.

With regard to what perceptual learning can tell us about the appropriate architecture for a model of speech perception, the intermediate levels (such as features and phonemes) are of particular interest because they are the most controversial. The key issue is at what point continuous speech information is categorized. What representations lie between continuous acoustic cues and words? Different models of speech perception posit categorization at different points. Some prototype models, like TRACE and MERGE (McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000), predict that categorization of acoustic input occurs at some intermediate level before words, as they have abstract units at a sub-lexical level. Not all prototype models may be as constrained: FLMP (Oden & Massaro, 1978) posits a continuous flow of information through various levels, with integration occurring at each level up to the decision point. While initially the syllable was proposed as the decision point, this claim is not central to the model. The decision point could be task dependent, and is likely to occur as late as possible (G. Oden, personal communication, October 3, 2011). Exemplar models do not predict any type of categorization before words, so they require no intermediate units (e.g. Goldinger, 1996, 1998; Johnson, 1997). By investigating whether perceptual learning generalizes or is specific at the levels of both words and phonemes, we can find out which intermediate representations are necessary and which are superfluous.

Using perceptual learning data to infer the types of levels and abstractions necessary for speech perception is not novel to this dissertation. This same approach has already been taken by McQueen et al. (2006); Cutler (2010); Cutler, Eisner, McQueen, and Norris (2010). One of the main points in all of these papers is that perceptual learning generalizes to untrained items (such as novel words), and that this type of generalization requires sub-lexical abstraction (such as phoneme units). The problem with this approach to date is that while there have been numerous experiments that tested generalization of perceptual learning, there have been very few experiments that tested specificity. This is unfortunate because specificity is equally critical when making strong claims about the necessity of abstract units in the perceptual system. Abstract units support and predict generalization, so generalization has been used as evidence in favor of abstraction. However, abstract units also predict that highly specific perceptual learning is not possible because of the information that is lost or discarded along the way. For example, if perceptual learning generalizes across words because listeners use abstract sub-lexical units like phonemes, then listeners should not be able to perceptually learn lexically specific boundaries. Similarly, if perceptual learning of a boundary between *voiced* and *voiceless* sounds relies on feature-level abstraction, then listeners should not be able to learn separate boundaries for the same feature in different phonemic contexts (such as /b/ and /p/ vs. /g/ and /k/). This is a simplification in that it assumes both a single-stream system and discarding of information at each level of abstraction, but models have made these simplifications as well. By only looking for generalization and not testing specificity, previous perceptual learning experiments have failed to make the stronger test of sub-lexical abstraction in speech perception. It is surprising that this avenue of investigation has been overlooked, especially given that the talker specificity of perceptual learning has not been so neglected. Both generalization and specificity are important for distinguishing models of speech perception, and the consideration of the specificity (in addition to generalization) at the level of words and phonemes is a novel contribution of this dissertation. In the next few sections I discuss what patterns of perceptual learning results might be observed at different levels in the speech perception system, and which models and theories would be consistent with each pattern of results.

1.1.1 The Lexical Level

Generalization of learned category boundaries across different words would be observed if listeners learned a shifted boundary between a pair of words like *park* and *bark*, and subsequently applied that boundary to two other words like *paste* and *baste* that were not present during training. This generalization across words would suggest that the speech perception system contains abstract sub-lexical units that listeners can use to generalize. For successful generalization, the boundary would need to be learned somewhere below the level of words, either at the level of the phonemes /b/ and /p/, the level of features (voiced and voiceless), or at some other abstract unit. MERGE and TRACE both predict this pattern of results because they have abstract sub-lexical units (McClelland & Elman, 1986; Norris et al., 2000).

In contrast, lexically specific learning would be observed if listeners learned a shifted boundary between *park* and *bark* but did not shift the boundary between *paste* and *baste*, or if they were able to simultaneously learn different boundaries for *park/bark* and *paste/baste*. This lexical specificity would not indicate a need for any type of abstraction before the level of words, suggesting instead that words may be the level at which speech information is categorized. This would mean that phoneme-like units, features, or other sub-lexical units are either unnecessary for speech perception or are not involved in perceptual learning for speech. It would furthermore indicate that acoustic cues can be mapped (and re-mapped) directly onto words. This pattern of results would be consistent with exemplar or episodic models, which lack sublexical units (Goldinger, 1996, 1998; Johnson, 1997).

To summarize, models of speech perception that have abstract sub-lexical units predict generalization of perceptually learned category boundaries across words. Models without such units predict lexically specific learning of category boundaries. It would be strange to encounter both generalization and specificity, which could be seen if participants were able to learn lexically specific boundaries but could also generalize across words. Since existing models either have or do not have sub-lexical units, it would be difficult for these models to accommodate this unexpected combination of results without some modification.

One possible hybrid model that could accommodate both lexical specificity and generalization is a dual path or triangle model, an example of which is shown in Figure 1.1. This type of model would allow listeners to use abstract sub-lexical representations like phonemes in order to generalize category boundaries across words, but also have an additional route directly to words (bypassing any sub-lexical units) that would allow for lexically specific boundary learning. This model is analogous to the triangle model for reading (Seidenberg & McClelland, 1989). In the triangle model for reading, readers can access words either directly from graphemes (sight-reading), or by way of phonemes (sounding out words). Sight-reading is a much faster and more direct path to meaning, but the phoneme route is better for some tasks, and may be all that learners have access

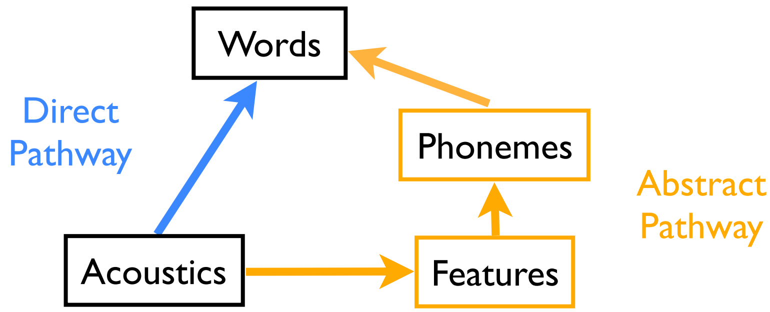

Figure 1.1: Dual-route or triangle model example. The direct pathway leads from acoustics to words without intermediate levels of abstraction. A second pathway leads from acoustics to words with different types of abstract units (such as features and phoneme) along the way.

to for other tasks (like reading unfamiliar words). A triangle model for speech perception might be similarly advantageous, allowing listeners to typically rely on a direct path to words, but also allowing them to use a path through phonemes or some other sub-lexical unit. This indirect path would presumably be advantageous for learning and performing meta-linguistic tasks.

The triangle model has an architecture that that shares some commonalities with MERGE, and with some modification MERGE could account for the same results (both lexical specificity and generalization) as well. In order to do generalization of learned category boundaries, MERGE would need connections from phonemes to words in addition to its existing connections from words to phonemes. These would essentially make MERGE into a version of the triangle model, or a hybrid MERGE-triangle model. However, bi-directional connections between phonemes and words would go against the nature of MERGE as a model without feedback.

The triangle model also shares some characteristics of the dual-stream speech processing model proposed by Hickok and Poeppel (2007). While both models offer an account of speech processing that involves multiple processing routes, one of which may be helpful for learning and speech production, the two pathways proposed do not appear to be equivalent. Hickok and Poeppel's model has a ventral stream pathway for mapping sound to meaning and a dorsal stream pathway for mapping sound to motor representations, while the triangle model has two routes to meaning. While the dorsal pathway appears to offer a secondary route to meaning as well, a second difference is that in Hickok and Poeppel's model, phonological processing occurs before the two pathways diverge. In our proposed triangle model for speech perception it is critical that only one of the two pathways involves this type of sub-lexical abstraction. This difference means that the Hickok and Poeppel model would not be able to account for both lexical specificity and generalization in perceptual learning, even though it has two processing routes, since neither of the routes provides a path to the lexicon independent from sub-lexical abstraction. This model does suggest, however, that a dual-path model of speech perception is not entirely unreasonable.

Adaptive Resonance Theory or ART (e.g. Grossberg, Boardman, & Cohen, 1997; Grossberg, 2003; Goldinger & Azuma, 2003) is a second class of model that might be able to handle both lexical specificity and generalization across words. In fact, ART appears to be able to handle any pattern of results (lexical specificity, generalization across words, and the combination of these results), although it is unclear whether this is true. ART does not specify the units that exist within the speech perception system: instead, the units are learned attractor states that can be nested, since speech input can be consistent with many different size units (e.g. phonemes, syllable, and words). During processing, the speech perception system will eventually reach a stable state where a particular unit achieves resonance. Task demands contribute to the relative weighting of top-down versus bottom-up information, which influences the size of the units that can achieve resonance. This flexibility, along with the lack of defined levels and direct connections between levels, allows ART to predict many different patterns of results based on the particular stimuli and task demands that a listener might encounter. Some situations will favor resonance and learning at the level of words, while others may favor sub-lexical units.

FLMP (Oden & Massaro, 1978) is another model that might be able to handle both lexical specificity and generalization across words if the decision point used by listeners is flexible. If listeners did not make voicing decisions until the word level, we should see lexically specific learning of category boundaries. An earlier decision point would lead to generalization across words. This presumes that some task demands would pressure listeners to either delay decisions in some circumstances, or make early decisions in other circumstances, and that we should see lexical specificity in the first case and generalization across words in the second.

Finally, parsing models like C-CuRE (McMurray & Jongman, 2011; Cole, Linebaugh, Munson, & McMurray, 2010; McMurray, Cole, & Munson, 2011) should also be able to accommodate a combination of lexical specificity and generalization. Listeners may store lexically-specific information and use this to conditionalize cue values on individual words or lexical contrasts, leading to lexically-specific boundaries. Generalization would be seen when listeners lack this lexically-specific information, or fail to conditionalize upon it, instead relying on global boundaries.

1.1.2 The Phoneme Level

While examining the lexical specificity and generalization of perceptual learning can tell us about whether models of speech perception require phonemes, looking at the phoneme level can tell us about whether the same models require features. As with the lexical level, generalization of learned category boundaries across different phonemic contexts that are distinguished by the same feature contrast would suggest that the perceptual system contains abstract feature units. For example, if a listener learns a new boundary for /b/ and /p/, generalization of that boundary shift to a coronal context (/d/ to /t/) would suggest that it was a voiced to voiceless feature boundary shift that occurred (rather than a phoneme-specific /b/ to /p/ boundary shift). TRACE and MERGE both predict this effect since they have abstract feature representations (McClelland & Elman, 1986; Norris et al., 2000).

In contrast, phoneme-specific learning would not require a level of abstraction before phonemes. If a learned boundary shift between /b/ and /p/ did not generalize to a different phonemic context like /d/ to /t/, this would be an example of phoneme-specific learning. This level of learning specificity would suggest that features are unnecessary or not important for perceptual learning in speech, and that acoustic cues are mapped to higher-level representations (either phonemes or words). Exemplar or episodic models of speech perception predict this pattern of results since they do not have abstract representations other than words (Goldinger, 1996, 1998; Johnson, 1997).

At the level of phonemes, it would again be surprising to encounter both specificity and generalization, and most models would find this combination of results difficult to account for. ART (Grossberg et al., 1997) is an exception since its lack of predefined units appears to allow it to account for any pattern of results: phoneme specificity, generalization across phonemic contexts, or both. The one pattern of results that ART could not account for is lexical specificity and phoneme generalization, but no other models could accommodate this pattern of results either.

A second model that could account for a combination of phoneme specificity and generalization would be a dual-path or triangle model like the one mentioned previously. To accommodate both types of phoneme results, it would be important to have a path directly from acoustic cues to phonemes or words (which would allow for specificity), plus a second path through features (which would allow for generalization). It is important to note that a single triangle model could handle the combination of generalization and specificity at both the lexical and phoneme levels, since the model needed to accommodate complex phoneme results would work equally well with multiple routes to phonemes or multiple routes to words.

Similarly, FLMP and C-CuRE (Oden & Massaro, 1978; Cole et al., 2010) could accommodate both phoneme specificity and generalization in the same way that they could handle lexical specificity and generalization. FLMP could do so by flexible movement of the decision point, and C-CuRE by conditionalizing (or not conditionalizing) upon phoneme-specific representations.

1.1.3 Talker Compensation

Another variant in the architecture of different models of speech perception is the point at which talker compensation occurs. Here again perceptual learning may be able to provide evidence to determine the appropriate architecture: where (or whether) a model of speech perception should incorporate talker compensation. If perceptual learning is talker-specific, it would indicate that the speech perception system maintains talker-specific representations or does not normalize for talkers, while perceptual learning that generalizes across talkers would suggest that the system does involve a process of talker compensation.

Generalization across talkers would occur if a perceptually learned boundary shift learned for one talker also shifted the boundary for a different talker. For example, if listeners were exposed to a female talker with a shifted boundary, and later tested on a male talker, they might generalize the shifted female boundary to the male talker as well. If this type of generalization occurred it would indicate that talker compensation must occur at or before the level of abstraction where the boundary is observed. If the boundary was at the phoneme level, this would mean that talker compensation occurs at or before phonemes. TRACE is an example of a model with this type of early talker compensation—talker-specific information is stripped out on the way to phoneme representations (McClelland & Elman, 1986). If the boundary was at the word level, it would mean that talker compensation must occur at or before words (in which case sub-lexical units could still be talker-specific).

The opposite effect, talker-specific learning, would occur if a boundary learned for one talker did not transfer to another talker. For example, if listeners learned a boundary shift for a female talker but when tested on a male talker showed no evidence of a boundary shift, it would suggest that their learning of the new boundary for the female talker was specific to that particular talker. Talker-specific learning would indicate that talker compensation does not occur until after the level at which boundaries are perceptually learned, or doesn't occur at all. This is the prediction made by exemplar models, which do not have talker compensation: talker-specific characteristics are preserved even at the level of the lexicon (Goldinger, 1996, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert, 2001; Pisoni, 1997).

A combination of talker-specific learning and generalization across talkers would not be quite as surprising as the combination of phoneme-specific boundary learning and generalization across phonemic contexts–exemplar models (e.g. Goldinger, 1996, 1998; Johnson, 1997) should be able to accommodate some degree of both generalization and specificity in this particular area, though their ability to show generalization is fairly limited. Since talker-specific information is preserved, generalization across talkers should depend on their acoustic similarity or experience with the most recently trained talker "overwriting" the previously trained talker. Talker-specific boundaries should be observed for talkers with fairly different voices (perhaps male and female talkers), while generalization might be observed for talkers with less distinct voices (such as two female talkers). Generalization might also be observed because of limited exemplar storage: if listeners can only store a limited number of exemplars, experience with a new talker might overwrite the exemplars from a previous talker, leading to generalization based on the more recently trained talker. The storage of talker-specific information in the lexicon could allow exemplar models to account for both talker-specific boundaries and generalization of boundaries across talkers without any extra processing for talker compensation, but only if generalization across talkers was limited to talkers with fairly similar voices or occurred only after blocked training. Exemplar models would also be unable to account for talker-specific boundaries and generalization across talkers occurring with the same two talkers, though other models may be able to account for this pattern of results.

The same dual-path model that could account for the combination of boundaries that could be specific to pairs of words or phonemic contexts, but also allow for generalization, would able to account for talker specificity and generalization as well. The path from acoustics to words, which allows for lexically specific boundaries, would also need to be talker-specific, while the second path through an abstract sub-lexical representation like phonemes or features would allow for generalization across words or phonemic contexts, and talkers.

C-CuRE could also accommodate both specificity and generalization. Listeners may track talker-specific characteristics, allowing them to conditionalize cue values based on talker-specific representations stored in the feedback connections between talkers and individual cues. This would also allow the model to use (and generalize) a single bottom-up mapping from cues to categories.

1.1.4 Theoretical Summary

The degree of *both* specificity and generalization found in perceptual learning for speech has important implications for theories and models of speech perception. Generalization would provide evidence in favor of abstract units like the phoneme and feature units in MERGE and TRACE. More specific learning would suggest a lack of abstraction, and would thus be consistent with models that do not have pre-lexical abstraction (like the episodic model). The combination of generalization and specificity at the same level would be unexpected and difficult for most models to accommodate, but might be compatible with ART, a dual-route model, C-CuRE, or FLMP. While the specificity and generalization of perceptual learning have not both been examined for words, phonemes, and talkers, there have been many studies looking at some of these issues. The next sections review this literature before and states the specific aims of the experiments that follow.

1.2 Basic Findings in Perceptual Learning

The majority of recent studies investigating perceptual learning in speech have used a paradigm first developed by Norris et al. (2003) in a study using Dutch fricatives. The paradigm depends on an effect known as the Ganong Effect, in which listeners are biased toward a word interpretation when listening to stimuli on a word to non-word continuum—their boundaries shift so that they make more word than non-word responses (Ganong, 1980). Listeners in the Norris et al. (2003) study were exposed to ambiguous fricatives (halfway between /f/ and /s/) spliced onto the ends of either /f/- or /s/-final words. These lexical contexts biased listeners to perceive the ambiguous sounds as either /f/ or /s/, depending on which context group they were in. For example, an ambiguous fricative spliced onto the end of “house” would likely be interpreted as an /s/ since “houf” is not a word; the same fricative spliced onto the end of “staff” would be heard as an /f/ since “stass” is not a word. Norris asked whether repeated exposure to such contexts would result in listeners shifting their category boundaries between the fricatives. Listeners who heard the ambiguous fricative in the /f/ context heard normal endings for the /s/ final words, and vice versa. Listeners performed a lexical decision task with these stimuli over several hundred trials, and then performed a fricative categorization task on sounds along an /ɛf/ to /ɛs/ continuum.[1]

During this subsequent categorization task, listeners showed boundary shifts that reflected their experimental group: those participants who heard ambiguous /f/ sounds showed a boundary closer to /s/, while those for whom the /s/ was ambiguous showed a boundary closer to /f/. In other words, the boundary shifted toward /s/ for listeners who heard the ambiguous fricative in /f/ final words, and toward /f/ for listeners who heard the ambiguous fricative in /s/ final words. Norris argued that listeners used their lexical knowledge to help them learn how they should interpret ambiguous speech sounds: feedback from the lexicon was used to modify some pre-lexical representation in order to better accommodate the ambiguous fricatives.

Sjerps and McQueen (2010) replicated this result using /θ/ instead of an ambiguous fricative between /f/ and /s/—listeners in the experiment were Dutch, so /θ/ did not map onto their native-language phoneme categories. This experiment showed that the Norris et al. (2003) results were not an artifact of the method used to create the ambiguous fricative, which involved blending recordings of naturally produced /f/ and /s/ fricatives. It is possible that during perceptual learning listeners were really just learning to ignore the irrelevant aspects of the blended fricative, and hear it as either the /f/ or /s/ depending on their training group. However, listeners in the Sjerps and McQueen (2010) study were trained to accept /θ/-final words as /f/ or /s/ final words, which means it is unlikely that the Norris et al. (2003) results were due to this type of selective filtering.

The same lexical feedback paradigm has since been used to show perceptual learning (as evidenced by shifted category boundaries) for a variety of phoneme contrasts. For fricatives it has been shown with /f/ and /s/ (Eisner & McQueen, 2005, 2006; Norris et al., 2003; McQueen et al., 2006; Sjerps & McQueen, 2010) and /s/ and /ʃ/ (Clarke-Davidson et al., 2008; Kraljic & Samuel, 2005, 2007; Kraljic et al., 2008). For stops

[1]The most ambiguous step along this continuum, as determined by stimulus piloting, was used as the ambiguous fricative at the end of the words in the lexical decision task.

it has been shown with /d/ and /t/ (Clarke & Luce, 2005; Kraljic & Samuel, 2006, 2007). Finally, with vowels it has been used for /i/ and /e/ (McQueen & Mitterer, 2005).

While the boundary shifts observed using this paradigm have been assumed to reflect changes in category representation, there was initially no evidence that they were not merely due to a change in decision bias following the training task. Clarke-Davidson et al. (2008) investigated this possibility with two experiments and a signal detection analysis. First they replicated the training and categorization tasks used by Norris et al. (2003) and added an AXB discrimination task. The discrimination task showed that the location of the peak in discrimination changes based on training and in accordance with the boundary shift, which suggests changes in category representations rather than decision bias. In a second experiment Clarke-Davidson et al. (2008) sought to reduce decision bias introduced by the training task, so they trained listeners using a same-different discrimination task rather than the lexical decision task. Both the boundary shift and discrimination peak results were replicated in this experiment. Finally, Clarke-Davidson et al. (2008) used a signal detection analysis to separate the behavioral effects due to decision bias from the effects due to category change. While this analysis indicated that the discrimination training task (meant to reduce decision bias) actually introduced a decision bias not seen with the lexical decision training task, boundary shifts were driven by both decision bias and category remapping, and changes in discrimination were driven solely by changing categories.

In addition to being driven (at least partially) by true changes in category representation, the boundary shifts produced with the lexical feedback paradigm are surprisingly persistent over time. Kraljic and Samuel (2005) found that shifts grew larger after a 25 minute delay, while Eisner and McQueen (2006) found that they lasted for 12 hours, even when listeners heard speech from other talkers during that time. However, under some circumstances, they can disappear quite rapidly. Kraljic and Samuel (2005) found that hearing good (unambiguous) tokens of the previously-ambiguous phoneme caused the shift to disappear, as long as these unambiguous tokens were produced by the same talker who produced the ambiguous tokens. The persistence of boundary shifts (across intervening talkers) and the finding that they are not disrupted by speech input from new talkers both suggest that perceptual learning might be fairly specific with regard to different sources of variability in the speech signal. They suggest talker-specificity in particular; however, none of these studies have directly tested this (e.g., trained listeners on two talkers with different boundaries). Moreover, learning may also be word or phoneme specific—this has not been extensively examined. This dissertation addresses issues of generalization and specificity in learning at the level of individual talkers, phonemes, and words. Each of these topics is reviewed in more depth below.

1.3 Word-Specific Learning

While no studies have addressed the possibility of independently adjusting phoneme category boundaries for different minimal pairs of words, six studies have examined how category boundaries generalize across words. As I shall describe, five found evidence that perceptual learning generalizes: Allen and Miller (2004); McQueen et al. (2006); Maye, Aslin, and Tanenhaus (2008); A. Hervais-Adelman, Davis, Johnsrude, and Carlyon (2008); Sjerps and McQueen (2010), and one did not (Buchholz, 2009). In addition to these studies there have been a number of experiments on generalization across phonemes. These experiments are discussed in the following section on phoneme-specific learning, but they are worth mentioning here as well since generalization across different phonemic contrasts relying on the same acoustic cues necessarily requires generalization across words. Of these studies, two found evidence of generalization (Kraljic & Samuel, 2006; Theodore & Miller, 2010) and three were inconclusive (Clarke & Luce, 2005; Maye, Weiss, & Aslin, 2008; McQueen & Mitterer, 2005). An additional pair of related studies, though not on the topic of phoneme generalization in particular, suggest that perceptual learning may occur at a level of abstraction below phonemes (Skoruppa & Peperkamp, 2011; A. G. Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011). This would prohibit both phoneme- and lexically-specific perceptual learning. Thus, a fair amount of evidence suggests that lexically-specific learning may not be possible, and that learning is constrained at the phonemic or sub-phonemic level of processing. Exemplar or episodic models, however, predict that lexically-specific learning should be possible (e.g. Goldinger, 1996, 1998; Johnson, 1997).

McQueen et al. (2006) tested this hypothesis by looking at whether perceptual learning for ambiguous fricatives affected the interpretation of words that were not used during the training or exposure phase of the experiments. If perceptual learning for the shifted phoneme category boundary was lexically specific, it should not have an effect on the perception of other words. Training was done with a lexical decision task as in Norris et al. (2003), where listeners heard an ambiguous fricative between /f/ and /s/ replacing either /s/ or /f/ word-finally, and the other fricative was produced normally. After training, listeners completed a cross-modal identity priming task where they heard an auditory prime and made a lexical decision task in response to a visually presented word or non-word. The auditory prime words ended with the ambiguous fricative used during training, and were minimal pairs that were words whether completed with an /f/ or an /s/. The boundary shift training was effective, as demonstrated by response time and word acceptance rate interactions between training condition and fricative ending. For cross-modal identity priming, there was a three-way interaction between prime type (ambiguous vs. unrelated), training condition, and target word ending (/f/ or /s/), which indicated that training condition affected the way that listeners responded to /f/ and /s/ words primed by ambiguous fricatives. Subsequent analyses showed that participants who were trained with ambiguous fricatives in /f/-final words were faster to respond to/f/-final targets when primed with an ambiguous fricative instead of an unrelated prime. The same pattern of results was seen for participants who were trained with the ambiguous fricative at the end of /s/-final words: they were faster to respond to /s/-final targets when primed with an ambiguous fricative than with an unrelated prime. The ambiguous fricative did not prime words that were trained with the natural endpoint fricatives for either training group. These priming results showed that perceptual learning affects the way that listeners respond to untrained words; that is, listeners generalize across words.

The lexical generalization result was replicated by Sjerps and McQueen (2010) using the same procedure as McQueen et al. (2006). In an additional experiment, /θ/ was used instead of the ambiguous /f/ and /s/ blend, so instead of hearing an ambiguous fricative at the end of either /f/ or /s/-final words, listeners heard /θ/. Since the listeners were Dutch, this sound did not map onto their native language phoneme categories. There was a difference between the training conditions on the lexical decision task, where words ending with /θ/ were more often accepted as words for the group that heard /θ/ in/f/-final words. On the cross-modal identity priming task used to test lexical generalization, hearing /θ/ at the end of prime words did prime either/f/ or /s/-final target words, depending on training groups. This provides further evidence that perceptual learning generalizes to untrained words.

Allen and Miller's (2004) study on talker-specific learning also included a word generalization test relevant to the question of lexical specificity in perceptual learning. Listeners were trained on two talkers with different voice-onset-times (VOTs) during the first session, and later returned for a second session where they were tested on new words that they hadn't heard during training. (VOT is a cue that distinguishes voiced and voiceless sounds like /b/ and /p/ or /d/ and /t/.)These generalization words shared the same onset consonant as the words from training. As in the test on the trained words, listeners were told which talker they were being tested on and selected which of two VOT variants of the test word was spoken by that talker. They showed generalization to the new words, selecting the VOT variants that were consistent with their training for each talker.

Maye, Aslin, and Tanenhaus (2008) also found generalization to untrained words in their study of perceptual learning for shifted vowels. During the lexical decision task, listeners responded to words they hadn't heard during a training task as well as words that they had heard during training. After training, listeners showed an increase in "word" responses to the vowel-shifted words, and this increase generalized to untrained words containing the same vowels.

Additional studies have examined generalization across words following training on different types of speech. While these studies involve perceptual learning, they do not focus on boundary shifts. A. Hervais-Adelman et al. (2008) looked at perceptual learning for noise-vocoded speech. On training trials their listeners heard a vocoded word, repeated it back, and then heard the word in its natural form and again in its vocoded form. They were trained on 120 words with 60 in each block. Across blocks the number of words and phonemes that they correctly produced showed significant improvement. Of interest here is that this improvement generalized to untrained words.

The one study that failed to find evidence of generalization in perceptual learning for speech was an experiment on adaptation to dysarthric speech (Buchholz, 2009). Here listeners exposed to dysarthric speech showed improvements in recognition accuracy for the original wordlist, indicating that learning had occurred, but performance on a list of novel items was similar to performance on the original items before familiarization had occurred. Buchholz hypothesized that dysarthric speech is more difficult to adapt to than typical synthetically manipulated speech, and suggested that more training might have resulted in better generalization to novel words.

As five out of six studies that tested generalization across words found evidence that listeners generalize to new words, this seems likely to be true. Perceptual learning is certainly of greater benefit to listeners if they can generalize learning to untrained words. However, the ability to generalize does not rule out the possibility of more specific learning. Although listeners seem to generalize perceptual learning across words, word-specific perceptual learning may still be possible if the task demands it.

Although the research on lexical specificity in perceptual learning has suggested that listeners generalize across words, there are no studies that have intentionally addressed the question of whether word-specific perceptual learning is possible. Research on lexical specificity for perception more generally, however, suggests that listeners do represent multiple variants of word forms in the lexicon. Some phonemes vary allophonically, meaning there are multiple acceptable ways of producing the same phoneme. A /t/, for example, can be produced canonically or it can be reduced (flapped) when following a stressed vowel. In conversational speech, talkers largely tend to produce the flapped form (Connine, 2004). Since /t/ has multiple allophones, listeners might store multiple variants of words that contain this phoneme (one with the canonical /t/ and the other with the flapped version).

Connine (2004) used a word to non-word continuum to study variant representation, examining the size of the boundary shift produced by the Ganong Effect—a bias to interpret stimuli as words instead of non-words (Ganong, 1980). Connine used the size of the boundary shift produced by the Ganong Effect as a measure of lexical activation. She found that the boundary shift is larger when the word end of the continuum contains the more frequent of two variant word forms (e.g. a flapped versus a canonical /t/). Similarly, for word forms that can be produced with intact or deleted schwas, the more frequent of the schwa forms leads to faster lexical decisions (Connine, Ranbom, & Patterson, 2008). Listeners are also more likely to judge the words as having three syllables when the form with the intact schwa is more frequent (Connine et al., 2008).

Finally, less frequent word forms produce small cross-modal priming effects (Ranbom, Connine, & Yudman, 2009).

The evidence that listeners respond differently to more frequent forms of words suggests that lexically-specific perceptual learning may be possible. That is, listeners must have learned that each specific word is associated with a specific phonetic instantiation. While this presumes learning, it does not attempt to manipulate it. Thus, we will test whether listeners can learn different boundaries for individual words that share the same CV at onset, despite their apparent ability to generalize perceptually learned boundaries across words.

1.4 Phoneme-Specific Learning

As with lexical generalization, there has been a fair amount of research on phoneme generalization in perceptual learning. After words, phonemes are the next obvious level of abstraction that listeners might generalize across on the basis of some lower level of abstraction, or conversely, the level of abstraction for which boundaries might be learned. If phonemes (or similar units of abstraction) are the critical unit for which perceptual learning of categories takes place, we would expect listeners to have category boundaries specific to individual pairs of phonemic contrasts. Whether or not listeners generalize across phonemes can tell us about what kinds of abstract representations they might have, and what role they play in speech perception.

Clarke and Luce (2005) exposed listeners to shifted or typical VOTs for /d/ and /t/ words in a word-monitoring task. The sentences used in this task were designed to contain only alveolar stops, and most of the words participants were monitoring for were not /d/- or /t/-initial. Participants also completed a categorization pre-test on syllables along a /ta/ to /da/ continuum, and then repeated this task every 20 sentences. They used a categorization task as a pre-test and then repeated it every 20 trials. Listeners in the shifted VOT condition showed corresponding boundary shifts during categorization, but when tested on a /g/ to /k/ continuum the shift did not clearly generalize. The control listeners showed a shift in the opposite direction for the /g/ to /k/ continuum while the experimental group did not, leading the authors to hypothesize that the lack of a shift might actually reflect generalization from learning a shifted boundary between /d/ and /t/. While this is possible, the argument is based on a null effect, so stronger evidence is needed to conclude that listeners generalize boundary shifts across phonemes. Although it is not critical for interpreting the phoneme generalization results of this experiment, it is important to note (as a methodological aside) that listeners in a two-task experiment like this could learn from both the categorization and word-monitoring tasks. The range of VOT exemplars that listeners are exposed to has been shown to affect their categorization boundaries (e.g. Rosen, 1979), so because the syllables in the categorization task used here had a different range of VOTs than listeners would typically experience while listening to English, this task could have contributed to the boundary shift seen for listeners in the shifted VOT condition. However, since there was no boundary shift observed for listeners exposed to typical VOTs during the word-monitoring task, the range of exemplars presented during this particular categorization task does not appear to be sufficient to drive a boundary shift on its own.

Returning to the topic of phoneme generalization, Kraljic and Samuel (2006) tested generalization across phonemes in addition to generalization across talkers. Listeners who initially heard ambiguous /d/ or /t/ sounds and were later tested on a /b/ to /p/ continuum showed generalization to this new continuum. This indicates that learning may be changing something about voicing in general, or how listeners shift the acoustic cues to voicing (e.g., VOT) for all phonemes, rather than shifting boundaries between specific phonemes. Most recently, the Theodore and Miller (2010) study on talker-specific learning also tested generalization across phonemes. Though listeners were trained only on /b/ and /p/ words, they were tested on both /p/ and /k/ words.

Since the talker-specific variation only affected the voiceless side of the training continuum, only words with voiceless onsets were used to test generalization to a new place of articulation. During testing listeners selected the VOT variant consistent with their training for each talker, for both the trained (/p/) words and the novel (/k/) words. This provides further evidence that perceptual learning may occur at the level of features (in this case voicing again), acoustic cues, or somewhere below phonemes. Theodore and Miller (2010) discuss the possibility that generalization could be based on raw acoustic similarity between trained and novel words.

If perceptual learning does adjust representations of acoustic cues rather than phonetic contrasts, then talker or lexical specificity may be very difficult to attain; rather generalization should be the norm. As a result, shifting one vowel contrast should affect all of the rest of the vowels as well. Here the evidence for generalization has been inconclusive. McQueen and Mitterer (2005) used the lexical feedback paradigm from Norris et al. (2003) to examine perceptual learning for shifted vowels. Listeners were exposed to words with /i/ and /e/ vowels with one of the endpoints replaced with an ambiguous vowel between /i/ and /e/. They found that listeners adapted to the vowel shifts, but generalization to untrained contrasts was weak. Maye, Aslin, and Tanenhaus (2008) also used lexical feedback to induce perceptual learning for vowels, but their method was different. They exposed listeners to vowels that had been shifted to an entirely different phoneme (/i/ pronounced as /ɪ/, rather than something ambiguous) in a passive listening task. Listeners heard a passage from the Wizard of Oz with the front vowels lowered. As in McQueen and Mitterer (2005), lexical context indicated the intended vowel. Listeners performed a lexical decision task before and after training. Following training they accepted more of the words with lowered vowels as real words, showing that they adapted to the vowel shift. Although there was some suggestion that the shift generalized to vowels that were not presented during exposure, the effect was not significant.

The results of these two vowel studies indicate that either perceptual learning does not generalize well across vowels, or that generalization effects are small and difficult to detect.

Although the previous two studies failed to find conclusive evidence for generalization across vowels, there is other vowel-based evidence that a feature-level representation may be important for perceptual learning. This would mean that we should observe generalization across phonemic contexts that rely on the same feature distinction, and we should not see phoneme-specific learning.

Skoruppa and Peperkamp (2011) found that French listeners adapt well to accents with consistent vowel harmony or disharmony, but not as well to an accent in which some vowels harmonize and other vowels disharmonize. This could be because listeners are making inferences on the level of phonological features or continuous acoustic cues, rather than phonemes. Like Theodore and Miller (2010), Skoruppa and Peperkamp (2011) point out that acoustic similarity could play the same role as features, so learning need not be based on abstract features per se.

However, there is other evidence that perceptual learning occurs not at the level of the raw acoustic signal, but after some abstraction has occurred. This evidence comes from a recent study on perceptual learning of vocoded speech that tested generalization to different frequency ranges. This study differs from the previously discussed studies in that the perceptual learning involved was not focused on phoneme boundary shifts, thus its relationship to the previous studies must be interpreted with caution. Here the dependent measure was how well listeners adapted to vocoded speech (such that it becomes more intelligible with increased exposure) rather than to shifted phoneme boundaries. A. G. Hervais-Adelman et al. (2011) trained listeners on either 20 hi-pass or lo-pass filtered vocoded speech sentences and tested them on 20 sentences in either the same or different frequency range. They found that listeners tested in the untrained frequency range were just as good at recognizing the final 20 sentences as the listeners who heard those sentences in the frequency range that matched their training. A second experiment tested generalization among three carrier signals used to create noise-vocoded speech (noise bands, pulse trains, and sine waves). Critically, these stimuli shared the same envelope cues but differed in their fine-structure cues. They found that listeners generalized partially but not completely across carrier signals. These two experiments suggest that perceptual learning does not occur at the level of the raw acoustic signal, but rather at some level of processing where there has been a degree of abstraction from the signal, though it is not clear whether this is at the level of continuous acoustic cues, features, phonemes or words. This abstraction allows generalization across frequency ranges, which could not have been entirely due to envelope cues since generalization in the second experiment was not complete.

Together the studies on generalization of perceptual learning across phonemes suggest that generalization can occur, though more research on this topic is needed. Two studies found evidence of generalization (Kraljic & Samuel, 2006; Theodore & Miller, 2010). Others suggest that generalization might occur, but failed to find concrete evidence that it does (Clarke & Luce, 2005; Maye, Weiss, & Aslin, 2008; McQueen & Mitterer, 2005). Such studies suggest that perceptual learning may be phoneme specific, which would mean that lower-level abstractions like features might not be necessary, at least for perceptual learning. In contrast, other studies on generalization of perceptual learning (though not about phonemes in particular) suggest that this learning may occur at the level of features or some other abstract unit below the level of phonemes (Skoruppa & Peperkamp, 2011; A. G. Hervais-Adelman et al., 2011). This would constrain perceptual learning in a way that would make it difficult for listeners to do phoneme-specific perceptual learning.

None of the studies on phoneme generalization addressed the question of whether it is possible to confine perceptual learning to a pair of phonemes (as opposed to learning a feature like voicing that applies to many phonemes), or to learn conflicting shifts for a given contrast in different phonemic contexts. Listeners doing perceptual learning may learn shifts for features or acoustic cues that they apply to all phonetic contrasts that rely on those features or cues. This would lead to generalization across phonemes. If this is the case, it would be difficult to shift different phoneme boundaries based on the same feature or cue in opposite directions. On the other hand, listeners do have different boundaries (for a single acoustic cue) for different phonemes—for instance, the VOT boundary between /d/ and /t/ is not the same as the boundary between /b/ and /p/—so this might make it possible for listeners to learn different shifts for different phoneme contrasts, even when those contrasts are based on the same acoustic cue. While models of speech perception do not differ in the predictions that they make about whether phoneme specific learning is possible (all predict that it is), there is little concrete evidence in favor of phoneme specific learning. Experiments on the specificity of perceptual learning for phonemes and the degree of generalization across phonemes may also help define the level at which perceptual learning occurs. I will look at phoneme-specific learning and generalization in Experiments 2 and 3.

1.5 Indexical/Talker-Specific Learning

Talker generalization, the inverse of specificity, is perhaps the best-studied domain in perceptual learning for speech. Initially, Eisner and McQueen (2005) tested whether a shifted boundary between /f/ and /s/ generalized to a novel talker. The stimuli used for testing in this study were unusual in that they cross-spliced vowels and fricatives from the trained and novel talkers, so when testing listeners on the "novel" talker, sometimes the fricative was actually from the trained talker (so only the vowel was produced by the novel talker). They found generalization to the novel talker (as determined by the vowel) only when the fricative was produced by the original (trained) talker, and they saw no generalization when both the vowel and fricative came from the novel talker. In their final experiment, however, listeners who were trained on the talker who had been the "novel" talker in the previous experiment did show effects of perceptual learning. This showed that the lack of generalization across talkers was not due to some characteristic of the novel talker's voice that might have prevented perceptual learning.

Kraljic and Samuel (2006) also examined talker generalization. They trained listeners on shifted /d/ and /t/ boundaries for either a male or female talker and then tested generalization with the talker of the opposite gender. Unlike the Eisner and McQueen (2005) study, here listeners showed generalization to completely novel talkers: the shifted boundary acquired under exposure to one talker generalized to a second talker. A subsequent study by Kraljic and Samuel (2007) addressed the question of whether boundary generalization differs for fricatives and stops, which could have caused the discrepancy between Eisner and McQueen (2005) and Kraljic and Samuel (2006). In this study, listeners were trained on two talkers, one after the other. In one experiment, listeners heard shifted stops (/d/ and /t/), and in another, other listeners heard fricatives (/s/ and /ʃ/). Training was blocked so that listeners completed training for one talker before they began training for the second talker. After training was completed, listeners performed speech categorization tasks for both talkers, one after the other. Listeners in the fricative experiment had talker-specific boundaries, while listeners in the stop consonant experiment generalized the most recently trained talker's boundary to the previous talker. Together, these studies on talker generalization support generalization of perceptually learned boundary shifts across talkers for stops, but not for fricatives.

Kraljic and Samuel (2007) suggest that the difference in talker generalization for stops and fricatives may indicate a difference in learning mechanisms for acoustic cues differentiated by spectral properties instead of temporal properties, but there are at least two alternative explanations for these results. The first is that stops and fricatives might not contain equal information about talkers, or that talker-specific information may not be as readily available in the stops. Indeed, phonetic analyses of fricatives suggest that talker variation is a significant contributor of the variance to almost all of the cues to fricatives (e.g., McMurray & Jongman, 2011; Newman, Clouse, & Burnham, 2001; Munson, McDonald, DeBoe, & White, 2006), while voicing (and VOT in particular), while still showing between-talker variability, may be more invariant with respect to talker (Allen, Miller, & DeSteno, 2003; Syrdal, 1996). That is, determining fricative place of articulation may require listeners to take into account talker identity in a way that stop-voicing may not. This could lead to greater generalization across talkers for stops. Secondly, and conversely, the overlap of talker-specific distributions may be greater for stop-voicing than for fricative place of articulation: talkers may vary more in their fricative productions than in their VOT productions. This could make it more difficult to determine talker identity based on stop-voicing than based on fricative place, thereby leading to greater generalization across talkers for stops than for fricatives.

In either case, however, while listeners in the stop consonant experiments showed generalization across talkers, these experiments primarily showed that generalization is possible, and were not designed to rule out talker-specific learning. Since talker-specific learning did appear to occur for fricatives, listeners may also be able to learn talker-specific boundaries for stop-voicing, but they may require greater exposure or a different training paradigm.

In contrast to these studies on generalization, two studies, both using a different paradigm, provide evidence for specificity, suggesting that it is possible to learn different stop-voicing category boundaries for different talkers. Allen and Miller (2004) and Theodore and Miller (2010) addressed the issue of talker specificity by testing whether listeners are capable of tracking individual talker differences on the basis of VOT. In Allen and Miller (2004), listeners were trained on two talkers, one with VOTs that were shorter than average for both voiced and voiceless words, and one with VOTs that were longer than average. Stimuli were synthetic speech with 3 different VOT variants for each talker. Of these variants, one was voiced (/d/) and two were voiceless (/t/). Each of the two voiced tokens was heard twice as often as the four voiceless tokens so that there was an even number of voiced and voiceless trials. The training task was a 4AFC task for which listeners selected both talker (Annie or Laura) and voicing category (/d/ or /t/). Feedback was provided for the talker decision but not the voicing decision. During testing, listeners were told which talker was being tested and decided which of two novel VOT variants sounded more like that talker. Listeners chose the test stimuli that corresponded with their training: they chose the shorter of the two test tokens for the talker with short VOTs, and the longer of the test tokens for the talker with long VOTs. Theodore and Miller (2010) replicated this result using the same design but with /b/ and /p/ words instead of /d/ and /t/. These two studies provide evidence that listeners are capable of learning different phoneme category boundaries for individual talkers.

However, because of the explicit emphasis placed on learning the two different talkers in the study, we do not know whether listeners automatically learn different phoneme boundaries for each talker. It is possible that they only do so when they receive prompting to pay attention to talker differences or when task demands require talker differentiation. Talker differences were especially emphasized in these studies since listeners received feedback on their talker decisions but not their VOT decisions. Additionally, both studies used synthetic speech, which does not have as many indexical cues as natural speech. It is possible that with other indexical cues being weaker than normal, listeners could be falling back on VOT as a cue to talker identity, when they might not typically pay attention to talker-specific variation in VOT. These experiments would then be an example of listeners learning to associate VOTs with talkers rather than talker-specific shifting of phoneme boundaries.

To summarize, Allen and Miller (2004) and Theodore and Miller (2010) found evidence of talker-specific learning while Kraljic and Samuel (2006, 2007) and Eisner and McQueen (2005) found generalization across talkers. These results are not as contradictory as they might seem, given that generalization across talkers does not really rule out talker-specific learning, but the apparent contradiction is also unsurprising given the many methodological differences between the studies. Both Allen and Miller (2004); Theodore and Miller (2010) emphasized talker identification and trained listeners on both talkers simultaneously, while listeners in the other studies were not told that talker identification was important and were trained on only one talker at a time (sequential talker training). It is possible that either of these differences in methodology could produce the discrepancy in results.

If simultaneous training does lead to talker-specific learning and blocked training leads to generalization across talkers, it would suggest that listeners need exposure to multiple talkers within a short timespan in order to form talker-specific representations. This effect could be explained in a number of different ways. One possible explanation is that listeners are unable to store speech information for an extended period of time. If listening to new speech overwrites storage of previously experienced speech, sequential talker training would lead to the most recent talker overwriting the previous talker. Another possibility is that intermixed trials of multiple talkers allow listeners to do some type of talker comparison that is not possible when they can't hear both talkers together. Finally, it may be that listeners only do talker-specific learning in situations where it appears that it may have some benefit. It may be more efficient to adapt a single set of boundaries when that appears to be sufficient for task performance (e.g. during single or sequential talker exposure), but when the task demands multiple boundaries (or lends itself to adjusting multiple boundaries), listeners are capable of talker-specific learning even though it is more challenging.

With respect to talker-specific issues in perceptual learning, this dissertation addresses a number of questions: 1) whether listeners spontaneously (or without prompting) learn talker-specific boundaries; 2) whether this learning occurs when the task does not require talker identification. If so, this would suggest a much more robust talker-specific learning mechanism than what is suggested by prior studies. Finally, I ask 3) whether simultaneous exposure to multiple talkers leads to talker-specific boundaries, while exposure to one talker at a time leads to generalization of boundaries across talkers. This allows us to examine the relationship of specificity to generalization. These questions will be addressed in Experiments 5 and 6.

Finally, if listeners do show evidence of talker-specific learning, and also show word-specific learning, then it would make sense to test talker by word-specific learning. While this would provide evidence for the most highly specific learning possible, it is unfortunately beyond the scope of this dissertation, and will thus remain an avenue open to further investigation.

1.6 Specific Aims

To summarize our current state of knowledge about generalization and specificity in perceptual learning for speech, most studies that have examined generalization to untrained lexical items have found that perceptual learning generalizes across words (Allen & Miller, 2004; McQueen et al., 2006; Maye, Aslin, & Tanenhaus, 2008; A. Hervais-Adelman et al., 2008; Sjerps & McQueen, 2010). However, while listeners may typically generalize across words, this does not necessarily preclude lexically-specific learning. Research on lexical representation of variant word forms suggests that listeners have multiple lexical representations for the same word (e.g. Connine, 2004; Connine et al., 2008; Ranbom et al., 2009), which suggests that lexically-specific boundary learning may be possible.

Results from studies on generalization of perceptual learning across phonemes have been more variable. While some have found evidence of generalization (Kraljic & Samuel, 2006; Theodore & Miller, 2010), other evidence has not been as strong (Clarke & Luce, 2005; Maye, Weiss, & Aslin, 2008; McQueen & Mitterer, 2005). Another line of research has suggested that perceptual learning occurs at the level of features (Skoruppa & Peperkamp, 2011; A. G. Hervais-Adelman et al., 2011), which would constrain perceptual learning such that phoneme-specific learning may not be possible. There has been no research directly addressing this question.

Studies on talker specificity and generalization using the lexical-feedback perceptual learning paradigm have suggested that listeners can learn talker-specific boundaries for fricatives but generalize across talkers for boundaries between stops (Eisner & McQueen, 2005; Kraljic & Samuel, 2006, 2007). However, other studies using a different paradigm have suggested that listeners are capable of learning talker-specific boundaries for stops as well (Allen & Miller, 2004; Theodore & Miller, 2010). Training differences between the studies may be responsible for the discrepancy in the results.

Despite the large amount of research that has been done on perceptual learning for speech, there are many questions that remain unanswered. With regard to words, we do not know if it is possible to learn different boundaries for words that share the same CV at onset. Similarly, with regard to phonemes, we do not know if it is possible to learn conflicting boundary shifts for a given feature contrast in different phonemic contexts. With regard to talkers, we do not know if listeners are able to learn talker-specific category boundaries without prompting, and under what circumstances this might occur. Previous studies have found generalization across talkers, perhaps due to blocking exposure by talker. These issues lead directly to the specific aims of this dissertation:

1) To test whether perceptual learning in speech is phoneme-specific. This aim will be addressed in Experiment 2, where continua with one onset phoneme will be shifted to the left and continua with another onset phoneme will be shifted to the right.

2) To assess the degree of generalization to different phonemic contexts that rely on the same feature contrast. This aim will be addressed in Experiment 3, where listeners will be exposed to shifted continua in a single phonemic context, and then tested on generalization to new words in the same and different phonemic contexts.

3) To test whether perceptual learning can be specific to particular words. This aim will be addressed in Experiment 4, where listeners will hear words with the same CV onset shifted in opposite directions.

4) To test whether spontaneous talker-specific perceptual learning can be observed task that does not emphasize talker identification. This aim will be addressed in Experiments 5 and 6, where listeners will hear two different talkers with speech distributions shifted in opposite directions.

5) To test whether sequential versus simultaneous talker training affects the degree of talker-specificity in learning. This aim will be addressed in Experiments 5 and 6 as well. Experiment 5 will use a mixed design with training on the two talkers interspersed. Experiment 6 will use a blocked design with training on each talker presented on a different day.

To meet these aims we needed a paradigm for perceptual learning. While the lexical-feedback paradigm used in many other studies would have been perfectly appropriate, we accidentally stumbled upon a different perceptual learning paradigm that could be used to shift categorization boundaries. Although we used this new paradigm, we could have used the lexical-feedback paradigm (e.g. Norris et al., 2003) instead. The next chapter explores these issues in more detail before presenting Experiment 1.

CHAPTER 2
METHODOLOGY AND PILOT DATA

2.1 A New Paradigm for Perceptual Learning

The majority of the studies reviewed in the introduction have relied on a type of implicit supervised learning to train listeners on shifted category boundaries. The difference between this and unsupervised learning is that supervised learning requires some kind of error signal. While these error signals are often thought of as a very explicit kind of feedback, they need not be. In the Norris et al. (2003) paradigm, lexical knowledge provides an error signal that helps fluent listeners interpret ambiguous speech sounds. Feedback from the lexicon biases listeners to perceive spoken language as words that they already know (Ganong, 1980). If a listener perceives an /s/ at the end of a string of sounds that only forms a word when ended with an /f/, lexical knowledge will help the listener figure out that the speaker probably meant to produce an /f/, and that they should subsequently remap this sound to their /f/ category. Cutler, McQueen, Butterfield, and Norris (2008) suggest that it is phonotactic knowledge rather than lexical knowledge that drives perceptual learning for shifted phoneme category boundaries, but this still allows for an error signal: simply one from a different source.

However, in many cases (e.g., learning a new language, minimal pairs where both forms are words) participants may not have access to this source of feedback. Here, unsupervised perceptual learning may be needed. In an unsupervised perceptual learning paradigm, listeners would shift their phoneme category boundaries without any kind of error signal telling them to do so, lexical or otherwise. One way this could occur is if listeners were sensitive to the distribution of sounds that they hear. For example, VOT typically shows two clusters centered around 0 and 50 ms with a boundary at 25 ms (e.g., Lisker & Abramson, 1964). However, if a talker's VOTs cluster around 15 and 65 ms, listeners might reasonably learn a new boundary at 35 or 40 ms. In this case, listeners hearing shifted distributions of speech sounds would shift their category boundaries to match the talker's distribution, changing their phonetic categories even in the absence of feedback from lexical or phonotactic knowledge. Indeed, Maye and Gerken (2000, 2001) have shown that adults are capable of extracting category structure (e.g., the number of categories) from a series of non-word stimuli based solely on their distributional statistics, and if listeners can determine the number of categories based on distributional statistics, then these same statistics might also help them determine the category boundary locations.

Unsupervised learning of phoneme category boundaries is interesting for a number of different reasons. The first of these is that distributional learning has been suggested as a likely mechanism for infant speech category development (Maye, Werker, & Gerken, 2002). Unsupervised learning is certainly a more plausible mechanism for infant speech category development than supervised learning, given that early in development, infants lack lexical or phonotactic knowledge as sources of error signals for supervised learning. Further, there is no reason to assume such a mechanism would not operate during adulthood, so adult category boundaries may remain sensitive to this same type of learning. Secondly, unsupervised learning would allow adult listeners to take advantage of information from all the words they are exposed to instead of only the ones that lack minimal pairs. When learning the boundary between /b/ and /p/, the words *beach* and *peach* would be uninformative for listeners relying supervised learning to bias them in one direction or another. However, these words are useful for unsupervised learning of /b/ and /p/ categories. Finally, supervised learning for speech relies on a process of inference: listeners hear an ambiguous sound and must infer its category membership based on the context in which it was heard, though they may not be aware of the process. Unsupervised learning for speech, in contrast, relies on the simpler process of data accumulation: categories may be defined based on how frequently different sounds occur. Supervised learning need not be dismissed as a useful mechanism because it relies on inference, but neither should a simpler mechanism be discounted when it may also contribute to learning.

If listeners are able to do both supervised and unsupervised learning, they should take advantage of all sources of information that are available to them. In fact, the perceptual learning effects observed in studies that have used the lexical feedback paradigm may be partially dependent on unsupervised learning. Listeners in these studies hear typical pronunciations for one speech category (e.g. /f/) and another category is entirely replaced by an ambiguous pronunciation for another (e.g. something between an /f/ and an /s/). In essence, the distribution has shifted so that the ambiguous sound has become the prototypical (most frequent) exemplar of a category. An unsupervised learning account of the boundary shift would suggest that it is this shift in statistical distribution that drives the categorization boundary shift, not lexical feedback. This does not appear to be true since no boundary shift is observed when the ambiguous sounds are embedded in non-words (Norris et al., 2003). However, in the lexical feedback condition (where ambiguous sounds are embedded in words), listeners could be using both supervised and unsupervised learning, while in the non-word condition the underlying statistics are the only possible source of learning. It could be the combination of effects due to both types of learning that allow us to observe a boundary shift in the word condition but not the non-word condition, and that unsupervised learning alone is not enough to drive the effect. This could be because unsupervised learning takes longer (more exposures) than supervised learning, or that effects of unsupervised learning are smaller and more difficult to detect.

Although most research on distributional or statistical learning has been done with infants (Maye et al., 2002; Maye, Weiss, & Aslin, 2008), there has also been work showing that adult listeners are sensitive to statistical information about the distributions of speech input. Clayards, Tanenhaus, Aslin, and Jacobs (2008) manipulated the shape of VOT distributions that listeners were exposed to and measured the effect that this had on their activation for lexical competitors. Listeners in this study had their eye-movements recorded while they listened to target words along VOT continua from /b/ to /p/ and clicked on corresponding target images. The VOT distributions were manipulated so that not all steps along the continuum were equally likely. Listeners who heard wide (high variance) distributions of VOTs made more eye-movements to competitor objects than listeners who heard narrow (low variance) distributions.

While this addressed the width of the categories—rather than their locations along the continuum (and the consequent boundary)—a version of this paradigm might be able to shift VOT boundaries as seen in the lexically driven studies on perceptual learning, but in an unsupervised perceptual learning task. If listeners track how often different VOTs occur, then they should shift their voicing category boundaries to the left or right based on the placement of the prototype steps (those that occur most frequently) along the VOT continuum. This should involve perceptual learning rather than selective adaptation because we are not manipulating the frequency with which listeners are exposed to the different endpoints of the continuum–both groups of listeners will hear a similar number of VOT exemplars associated with each voicing category (although this may vary slightly according to individual differences in voicing boundary location). As in Norris (2003), it is simply the side of the continuum that is ambiguous (either voiced or voiceless) that should drive perceptual learning, albeit through a different mechanism.

Experiment 1 examines whether listeners shift their VOT boundaries to correspond with the distributions they are exposed to in this unsupervised perceptual learning paradigm. Critically, in this paradigm, there is no lexical information for listeners to use when determining whether a given sound is voiced or voiceless, since there are words at both the voiced and voiceless ends of the continuum.

2.1.1 Eye-Tracking

Although Clayards et al. (2008) recorded eye-movements in order to measure activation of lexical competitors, our interest is in assessing phoneme category boundary shifts. This can be accomplished simply by recording mouse-clicks, so tracking eye-movements during these studies is not strictly necessary. However, eye-movements may provide a more sensitive measure than mouse-clicks alone. For example, eye-movements could allow us to address when (during real-time processing) different effects occur (e.g., McMurray, Clayards, Tanenhaus, & Aslin, 2008). Since eye-movements are sensitive to the timecourse of lexical activation, eye-tracking is well-suited to studying the timecourse of different effects. For instance, if listeners show evidence of talker-specific learning, we might ask if this effect is apparent at the earliest moments of lexical processing, or only after some initial processing that is talker-independent. It is possible that talker identity can only affect voicing judgments after some initial processing has occurred. Similar questions can be asked with regard to phoneme or word identification. Moreover, eye-movement analyses may also allow comparison of the timing of different effects across experiments.

2.2 Experiment 1

Experiment 1 establishes an unsupervised learning paradigm that can be used to shift speech category boundaries through perceptual learning. As a kind of pilot experiment for the paradigm this experiment does not have strong theoretical implications, and was originally run for a different purpose all together. Our discovery that perceptual learning of category boundaries can occur in this unsupervised manner was a serendipitous finding, and we present the results of this experiment here to simply demonstrate that listeners are sensitive to distributional statistics in a short-term perceptual learning study. Each listener was exposed to either a right- or left-shifted VOT distribution during the experiment. Since listeners did not hear multiple distributions or talkers, the experiment did not address questions of specificity or generalization and cannot be used to test learning specificity.

The design of the study is based on Clayards et al. (2008), who manipulated the shape of VOT distributions that listeners heard. For perceptual learning of category boundaries, our interest is in manipulating the location of the distributions rather than their shape. A study by Sumner (2011) suggests that moving the entire distribution—rather than the prototype steps alone—should be more likely to elicit a perceptually learned boundary shift. Sumner (2011) found that listeners trained on a speaker with a French accent showed a boundary shift only when trained on stimuli with variable VOTs. Though she used a perceptual learning paradigm based on lexical feedback, as in Norris et al. (2003), these results suggest that variability in VOT is critical for perceptual learning of shifted category boundaries.

This experiment involves a fairly large number of trials, which increased the likelihood of that we would be able to detect learning. Since the number of trials was large enough that it would be impractical or unpleasant for participants to complete the entire experiment in one day, participants completed the study in two sessions. Sessions were scheduled one week apart to make scheduling easier for participants, although they were able to reschedule follow-up sessions if they were unable to return to the lab at the same time a week after their initial session. Fortuitously, having sessions spaced a week apart also provides an opportunity to test boundary shift maintenance over a longer period of time than has been examined in previous studies, although this opportunity should only arise if listeners show evidence of a boundary shift by the end of the first session.

Table 2.1: Experiment 1 stimulus items.

/b/	/p/	/l/	/r/
beach	peach	lace	race
bees	peas	lake	rake
beak	peak	lei	ray
bit	pit	lock	rock
bin	pin	lamp	ramp
bill	pill	lane	rain

2.2.1 Method

2.2.1.1 Design

The first experiment tested whether shifting the prototypes of VOT distributions is enough to induce perceptual learning of shifted category boundaries. Half of the participants (the left-shifted group) heard categories centered at steps -10 and 40ms, and the other half of the participants (the right-shifted group) heard categories centered at steps 10 and 60ms. The exact distributions of VOTs within each category were roughly Gaussian and are shown in Figure 2.1. On each day of the two-day study, listeners heard 300 critical trials that consisted of words along six VOT continua. The specific tokens were selected according to the distribution assigned to each group. There were 50 experimental trials from each continuum, but the VOT distribution for each group was maintained across the six continua and not within each individual continuum. In addition to these 300 experimental trials, listeners also heard 300 trials of filler words with /l/ and /r/ onsets. As with the experimental trials, each filler pair was heard 50 times (25 times for each filler word). The words from the six continua and the filler items that they were paired with are shown in Table 2.1.

Table 2.2: Experiment 1 VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	4	36	70	36	4	4	36	70	36	4	0	0
Right-Shifted Distribution	0	0	4	36	70	36	4	4	36	70	36	4

Note: The prototype steps (the VOT steps with the most frequently occurring exemplars) were shifted two steps between distributions.

Along the VOT distribution used for the experimental trials, the two prototypical VOTs occurred most frequently and the VOT steps farthest from the prototypes occurred the least frequently. The two distributions used in the experiment are shown in Figure 2.1, which includes dashed lines marking the ideal boundaries for each distribution. Table 2.2 also contains the number of tokens heard at each step for each of the two distributions. These distributions were chosen to match the shape of the distributions used by Clayards et al. (2008) as closely as possible. The 300 experimental and 300 filler trials totaled 600 trials per day and 1200 trials across both days. As the planned talker-specific learning experiments were designed to have participants learn different boundaries for each of two voices, it was necessary to ensure that boundary shifts would work for both voices. Thus, half the participants in the current experiment were run with a male talker and half were run with a female talker. Talker condition was crossed with the left-/right-shift condition for a complete 2x2 design.

2.2.1.2 Participants

Participants were 38 individuals from the University of Iowa community who participated in the study in exchange for course credit or a nominal payment. All were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with university and APA standards. We had difficulty retaining participants across both days of testing, and

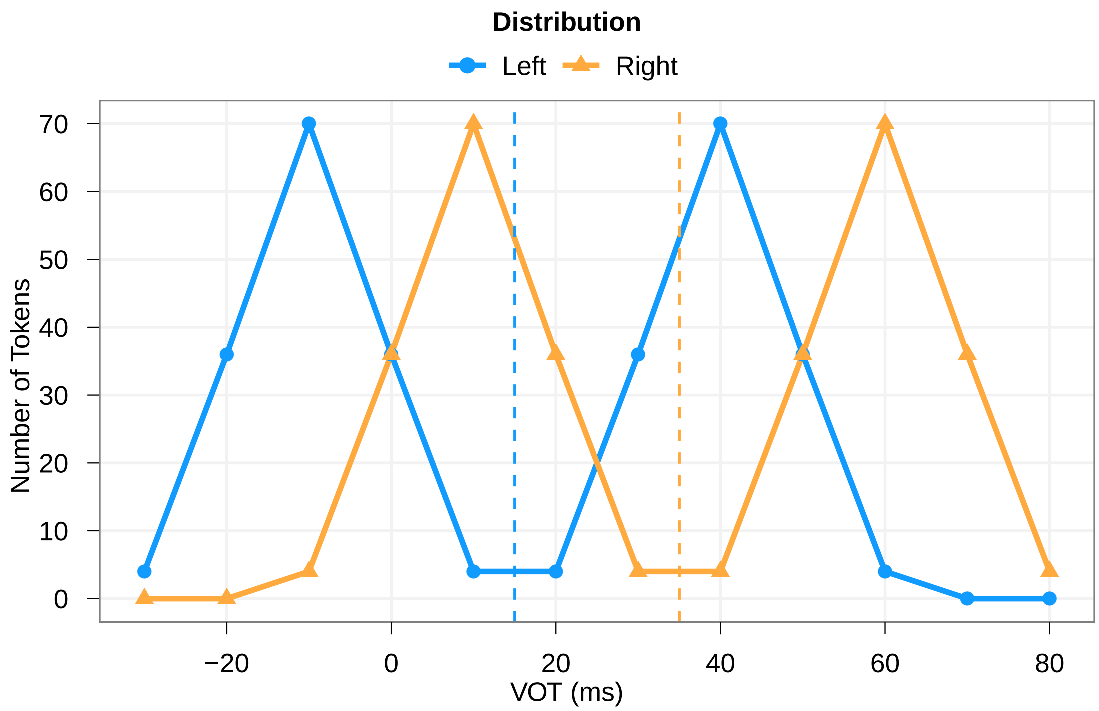

Figure 2.1: Experiment 1 VOT distributions, or the number of tokens heard at each VOT step for the left- and right-shifted distributions. The dashed lines at 15 and 35 ms indicate the ideal boundary locations for each distribution.

as a result approximately half of the participants did not return for the second day of the study. These were excluded from analysis, leaving a total of 17 participants who completed both days of the study.

2.2.1.3 Stimuli

2.2.1.3.1 Auditory Stimuli

Auditory stimuli consisted of twelve VOT continua (six per talker) ranging from -30 to 80 ms in twelve steps. Continua were created by cross-splicing recordings of natural speech using a technique similar to McMurray et al. (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008). First, both talkers were recorded in a sound-attenuated room using a Kay CSL 4501. Recordings were made using Praat sampling at 44.1 kHz. Talkers recorded several tokens of each word and we selected the best quality recordings with sufficient voicing and prevoicing for the experiment. The twelve filler words (six per talker) were recorded in the same recording sessions.

Next, we constructed the twelve continua from these recordings. Since each continuum contained both prevoiced and aspirated tokens, the prevoiced and aspirated portions were created using separate procedures. For both the prevoiced and aspirated sets, a voiced /b/ onset stimulus was selected (from the recorded tokens) to use as the base in both sets. A second /b/-initial word with prevoicing and a /p/-initial word with aspiration were selected to use for cross-splicing.[1] For the prevoiced portion of each continuum, progressively longer segments of prevoicing (beginning at the onset of voicing) were spliced onto the beginning of base stimulus. For the aspirated portion of each continuum, progressively longer segments of aspiration (beginning with the burst) replaced the onset portions of the base stimulus. As a result of this procedure, the vocalic segments were the same across the prevoicing and aspiration portion of each continuum (they were from the base stimulus). Aspiration and prevoicing were manipulated in approximately 10 ms increments, but because splicing was done at zero-crossings, splice points are not exactly 10 ms from each other. This is especially true for the prevoiced stimuli because there were fewer available zero-crossings in the prevoiced segments than in the aspirated segments. The VOT measurements of the completed stimuli are shown in Table 2.3.

Before splicing, each of the selected recordings had noise removed using Goldwave and was amplitude normalized in Praat. All recordings from the female talker were

[1]A second /b/ initial word was used for cross-splicing (instead of the same /b/ initial recording used as the base stimulus) so that neither the aspirated nor the pre-voiced sounds would be identity splices. Using the same /b/ onset stimulus for the pre-voicing portion and the base may have resulted in the prevoiced steps of the continuum sounding more natural than the aspirated steps, which we wished to avoid.

Table 2.3: Experiment 1 VOT measurements.

	beach/peach	beak/peak	bees/peas	bill/pill	bin/pin	bit/pit
Female Talker	-35	-34	-36	-31	-31	-31
	-23	-23	-24	-19	-25	-24
	-12	-11	-12	-12	-12	-12
	0	1	0	0	0	2
	10	10	12	10	10	11
	20	20	21	20	20	21
	30	30	30	30	30	31
	40	40	40	40	40	41
	50	50	50	50	50	51
	60	61	61	60	60	61
	70	71	71	70	70	71
	80	80	81	80	80	80
Male Talker	-32	-30	-32	-33	-29	-33
	-23	-17	-24	-25	-19	-23
	-12	-8	-8	-8	-10	-14
	0	0	0	0	0	0
	10	10	10	10	9	8
	21	20	20	21	20	20
	31	33	30	31	30	30
	41	40	40	41	40	40
	51	52	50	51	50	51
	61	60	60	61	60	61
	71	71	70	71	71	71
	80	80	80	81	80	81

scaled to maximum amplitudes of .95. Spliced stimuli from the male talker sounded over-aspirated when scaled to the same amplitude, so the base stimuli recordings were scaled to .99 and the aspirated recordings were scaled to .30. A few of the recordings required additional pre-processing, such as deleting a large initial peak in aspiration, duplicating short segments of aspiration to lengthen the aspirated portion of the word, or removing an audible click by deleting a pitch pulse.

Stimuli were piloted in a categorization task performed by six lab members. This piloting was done to ensure that all continua had boundaries at relatively central steps in the VOT continuum, and more importantly, that the cross-splicing manipulations still led to well-perceived endpoints that were consistently identified as either voiced or voiceless. On each pilot trial participants used a key-press to identify the given stimulus as beginning with /b/ or /p/. Each of the 144 stimulus items was repeated three times for a total of 432 trials per participant. At the /b/ endpoint, pilot listeners correctly identified the stimulus as a /b/ on 98.9% of the trials for the male talker, varying from 91.7% to 100% on the six different continua. They were at 97.3% correct for the female talker, from 84.6% to100% for the six continua. At the /p/ endpoint listeners were 100% correct for the male talker on all continua, and at 98.7% for the female talker (from 91.6% to 100% on individual continua). The point at which the identification functions for each talker crossed 50% was between steps 6 and 7 for both talkers, which is close to the middle of the 12-step continua.

2.2.1.3.2 Visual Stimuli

Pictures representing each item listed in Table 2.1 were constructed using a picture norming process that was developed in the McMurray lab to ensure that pictures were prototypical exemplars of their corresponding words (e.g., McMurray, Samelson, Lee, & Tomblin, 2010; Apfelbaum & McMurray, 2011). First, a number of candidate images were downloaded from a commercial clipart database. A committee of under-graduate and graduate students then selected the best (i.e. most prototypical) image. The committee also gave suggestions on how the selected images might be improved by changing colors, deleting unnecessary components, or adding in additional details. Finally, all pictures were edited to ensure uniform size and brightness. When appropriate images were not available in the commercial database, Adobe Illustrator was used to draw a clipart style image based on a reference photograph. Only the “ramp” image needed to be constructed this way. The final images were approved by the author and thesis supervisor.

2.2.1.4 Procedure

An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial of the experiment, participants were presented with four images from an item-set (a /b/, /p/, /l/, and /r/ onset item), one in each corner, and a red dot in the middle of the screen. After 500 ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial.

The 500 ms delay before the dot turned blue was included to give participants a chance to see what images were on the screen and where they were located before hearing the auditory stimulus. The sudden change in dot color was also likely to drive an eye-movement that would center eye-gaze at the onset of the auditory stimulus.

Throughout the experiment, participants' eye-movements were recorded by the Eye-Link II sampling gaze position every 4ms. Recording began at trial onset and lasted until a picture was selected. The Eye-Link II compensates for head movements so the participants were able to move freely during the study. Gaze position for both eyes was recorded when possible, but when calibration was not good for both eyes then one eye was tracked. When both eyes were tracked the one with better calibration was used during analysis. Drift correct events (to compensate for slippage of the eye-tracker on the head) occurred every 30 trials. Eye-tracking data was automatically parsed into saccade, fixation, and blink parameter based on the system's default parameters. Because the mouse-click results showed a fairly robust effect, we did not analyze the eye-movement data that was collected.

2.2.2 Results

First, we address overall task performance to establish that participants were paying attention to the auditory stimuli and could reliably identify the endpoints of the continua. The second section explains the analytic strategy and predicted results for the following two sections. The third section addresses perceptual learning of the shifted distributions, to show that group differences in boundary location emerged over the course of the experiment. The final section addresses talker differences, exploring whether the effect of distribution varied by talker.

2.2.2.1 Overall Performance

Mouse-click responses were first examined to assess overall performance on the task. On experimental trials where the stimulus began with a /b/ or /p/, participants clicked on a filler item on only 0.14% of the trials, which indicates that they were paying attention to both the auditory and visual stimuli.

Next, performance on the unambiguous endpoints of the continua was assessed. VOT steps -30, -20, and -10 were considered clear /b/ endpoints and steps 60, 70, and 80 were considered clear /p/ endpoints. Performance on both endpoints was very good. On the /b/ side, participants selected the /b/ image for 99.9% of the trials. On the /p/ side, participants selected the /p/ image for 97.8% of the trials. Endpoint performance, shown in Table 2.4, remained uniformly high when broken down by talker and distribu-

Table 2.4: Experiment 1 percent correct on endpoints by talker and distribution.

	Male Talker		Female Talker	
Endpoint	Left	Right	Left	Right
/b/	99.9	100	99.9	100
/p/	100	98.9	93.8	98.6

tion shift.

2.2.2.2 Analytic Strategy and Predictions

After establishing that endpoint performance was good, mouse-clicks across the entire continua were assessed to determine whether phoneme category boundaries differed between participants in the left- and right-shifted distribution groups, whether this difference increased over time, and whether participants learned shifted boundaries equally well for both the male and female talkers.

Both distribution and talker analyses were conducted with mixed effects models using the lme4 package (Version 2.12) in R (Bates, 2005) and a binomial linking function appropriate for binary data.[2] Mixed-effects models are advantageous for designs with both within- and between- participant effects, and for unbalanced designs, which makes this analytic strategy particularly appropriate for our experiments. While the number of data points in each distribution should be equal across the two distribution

[2]Technically our data are not binary because participants performed a 4AFC task, selecting the target image from the four images displayed on the screen. We used the binomial linking function as there is no multinomial linking function available, and this seemed like the most appropriate way to analyze our data. Although this is a simplification, treating our data as binary is appropriate because two of the response choices available were filler items. The fillers had very different names from the experimental items so they could be ignored on the experimental trials that we analyzed. This makes our task essentially a 2AFC task. We also excluded from analysis any trials for which participants selected one of the two filler images instead of one of the experimental images.

conditions, the two conditions have a different number of observations at each step. We chose not to use a curvefitting approach with these data because it would not be possible to get very good fits once we had binned the data (e.g. by participant, day, and continuum). The effects of learning and talker were assessed in separate analyses because of the number of factors necessary for each model, which led to over-specification when a combined analysis was attempted.

We initially examined a range of models to determine the best way to handle random-effects. Fixed effects were only examined after model selection. The initial models for each analysis included all of the fixed-factors and random intercepts for both random-effects factors. A second version of the models excluded the random effect of continuum, a third version included random slopes for participants, and a fourth included random slopes for both participants and words. Some of these models were over-specified and did not converge. Models were excluded if they did not converge within 300 iterations or were reported as having fitted probabilities of 0 or 1. The remaining models were then compared using Chi-Squared tests of model fit to select the best model.

For both analyses, if participants were learning category boundaries based on the shifted distributions, we predicted a main effect of distribution-group such that left-shifted group should be biased to respond with more /p/'s and the right-shifted group should be biased toward more /b/'s. For the learning analysis, we also expected to see a distribution × trial interaction showing that the group difference was larger later on in the experiment. Conversely, in the talker analyses, if both talkers support perceptual learning, we should see no interaction of talker and distribution.

2.2.2.3 *Learning Effects*

The first models looked at the effect of distribution over the course of the experiment. Response (/b/ or /p/, dummy-coded as 0 or 1 respectively) was the dependent variable. Distribution (left or right), day of the experiment (first or second), and half of the experiment (for each day) were all used as fixed-factors with two dummy-coded and centered levels. The VOT step of the stimulus was also treated as a fixed-factor. This was coded as a continuous covariate and centered. Only the eight middle steps of the continua (-10 to 60 ms) were analyzed as those steps were shared in the distributions for both conditions. Participant and continuum were treated as random-effects.

The random slopes models failed to converge so only the two random intercepts models were compared. The model with random intercepts for both subject and word was better than the model that excluded word ($\chi^2(2)=16.19$, $p<.001$), and was selected as the final model in the learning analysis.

The selected learning effects model is reported in Table 2.5. Critically, there was a significant main effect of distribution ($\beta=-1.85$, $p<.01$) indicating that participants in the left- and right-shifted distribution groups had different voicing category boundaries. There was also an effect of VOT step ($\beta=.22$, $p<.0001$), indicating that VOT affects voicing judgements. We would expect this to be significant in all of our models. The interaction between distribution and day ($\beta=-.51$, $p<.05$) was significant, suggesting that the observed boundary shift difference is due to learning over the course of the experiment since the effect is larger on the second day. This is illustrated in Figure 2.2, where each of the four panels shows a different quarter of the experiment. Initially there is little difference between the two distributions, but the difference grows as participants learn over time. By the end, the listeners who heard the left distribution have a categorization function that is shifted to the left relative to listeners who heard the right distribution. Finally, there was a three-way distribution by step by half interaction ($\beta=.05$, $p<.05$), indicating that the slope of the identification curves (as a function of VOT) changed over the course of the experiment. We did not conduct follow-up analyses on any slope effects because our interest is in the boundary differences, and in future models we do not

Table 2.5: Experiment 1 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.18	0.34	0.5	>0.6
Distribution	−1.85	0.63	−2.9	**<.01**
Day	−0.22	0.13	−1.8	>0.1
Half	−0.08	0.13	−0.6	>0.5
VOT Step	0.22	0.01	35.9	**<.0001**
Distribution x Day	−0.51	0.25	−2.0	**<.05**
Distribution x Half	0.23	0.25	0.9	>0.4
Day x Half	0.13	0.26	0.5	>0.6
Distribution x Step	0.01	0.01	0.5	>0.6
Day x Step	−0.02	0.01	−1.5	>0.1
Half x Step	−0.01	0.01	−1.0	>0.3
Distribution x Day x Half	−0.23	0.51	−0.4	>0.7
Distribution x Day x Step	0.01	0.02	0.5	>0.6
Distribution x Half x Step	0.05	0.02	2.3	**<.05**
Day x Half x Step	−0.02	0.02	−1.0	>0.3
Distribution x Day x Half x Step	−0.04	0.04	−1.0	>0.3

Note: The maximum correlation among fixed effects was r=.055 between distribution and step.

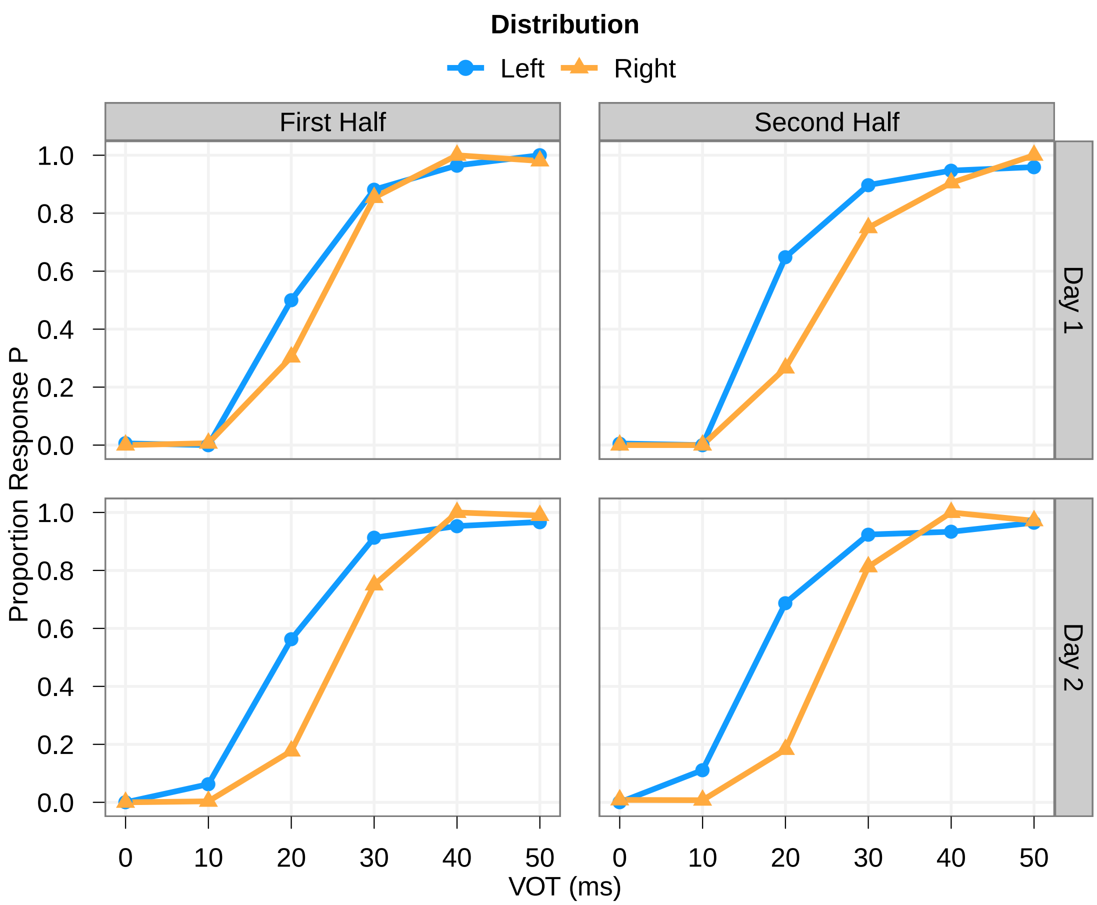

Figure 2.2: Experiment 1 distribution effect in each quarter of the experiment.

report VOT and slope effects in the text (although the relevant statistics will always be reported in a table).

Since the main analysis revealed an interaction between distribution and day, our next step was to run simple effects analyses to examine this in more detail. We wanted to determine when in the experiment an effect of distribution could be observed, as this tells us how rapidly participants learned the distributions and if learning persisted across the week between experiment sessions.

To run the simple effects analysis for each experiment session we split the data by day and applied the previous model (without day as factor) to each subset of the data. These two models are reported in Table 2.6. Both models showed effects of distribution (Day 1: β=-1.53, p<.02; Day 2: β=-2.22, p<.002). The significance of the distribution effect on the first day suggests that participants learned the distributions quite rapidly. Moreover, the larger boundary shift on the second day suggests that listeners retained what they had learned on the first day across the intervening week—if they had not retained what they had learned, they would have been effectively starting over again on the second day, and the boundary shift would have been the same on the two days.

2.2.2.4 Talker Effects

Since we used two talkers, one male and one female, we wanted to determine whether there was an effect of talker and if the effect of distribution varied by talker. That is, were participants able to learn the shifted distributions equally well for the two talkers? This analysis is similar to the prior analysis, but as the prior analysis showed similar learning effects across experiment days and the two halves of each day, we collapsed across the day and half factors to simplify the new model. Talker was dummy-coded (0 for male and 1 for female) and centered. A model with random slopes for both participants and words failed to converge and we compared three other models. The version with random slopes for participants and random intercepts for words was better than

Table 2.6: Experiment 1 simple effects for days one and two.

		Coef β	SE(β)	**z**	**p**
Intercept	0.24	0.34	0.7	>0.5	
	Distribution	−1.54	0.64	−2.4	**<.05**
	Half	−0.15	0.18	−0.8	>0.4
Day 1	VOT Step	0.23	0.01	24.4	**<.0001**
	Distribution x Half	0.30	0.37	0.8	>0.4
	Distribution x Step	0.01	0.02	0.6	>0.6
	Half x Step	0.00	0.02	−0.1	>0.9
	Distribution x Half x Step	0.07	0.03	2.2	**<.05**
	Intercept	0.11	0.40	0.3	>0.8
	Distribution	−2.23	0.69	−3.2	**<.01**
	Half	−0.02	0.18	−0.1	>0.9
Day 2	VOT Step	0.22	0.01	26.3	**<.0001**
	Distribution x Half	0.09	0.36	0.2	>0.8
	Distribution x Step	0.00	0.02	0.2	>0.9
	Half x Step	−0.02	0.01	−1.8	>0.1
	Distribution x Half x Step	0.02	0.03	0.9	>0.4

Note: The maximum correlation between factors was r=.059 in the day 1 model (between half and step), and r=-.042 in the day 2 model (between distribution and step).

Table 2.7: Experiment 1 talker model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.20	0.32	0.6	>0.5
Distribution	−1.85	0.59	−3.1	**<.01**
Talker	−0.69	0.59	−1.2	>0.2
VOT Step	0.22	0.01	34.5	**<.0001**
Distribution x Talker	0.19	1.19	0.2	>0.9
Distribution x Step	0.00	0.01	−0.3	>0.8
Talker x Step	−0.04	0.01	−3.1	**<.01**
Distribution x Talker x Step	0.07	0.03	2.8	**<.01**

Note: The maximum correlation among the fixed factors was correlations r=.34.

the next best model (χ^2(2)=159.075, p<.0001), so we selected this model.

The talker model for Experiment 1 is reported in Table 2.7. There was a significant effect of distribution (β=-1.85, p<.01) which observed previously in the learning analysis. Based on a plot of the data (Figure 2.3), we expected to find a distribution by talker interaction showing that the boundary effect was larger for the female talker, but this interaction was not significant in the model (β=.19, p>.8). There were, however, other interactions indicating that the slope effects were different across talkers. These were an interaction between talker and step (β=-.04, p<.01) and a three-way distribution by talker by step interaction (β=.07, p<.01).

2.2.3 Discussion

The first experiment established an unsupervised learning paradigm that can be used to shift category boundaries by perceptual learning. Listeners exposed to the left distribution had different (left-shifted) category boundaries than listeners exposed to

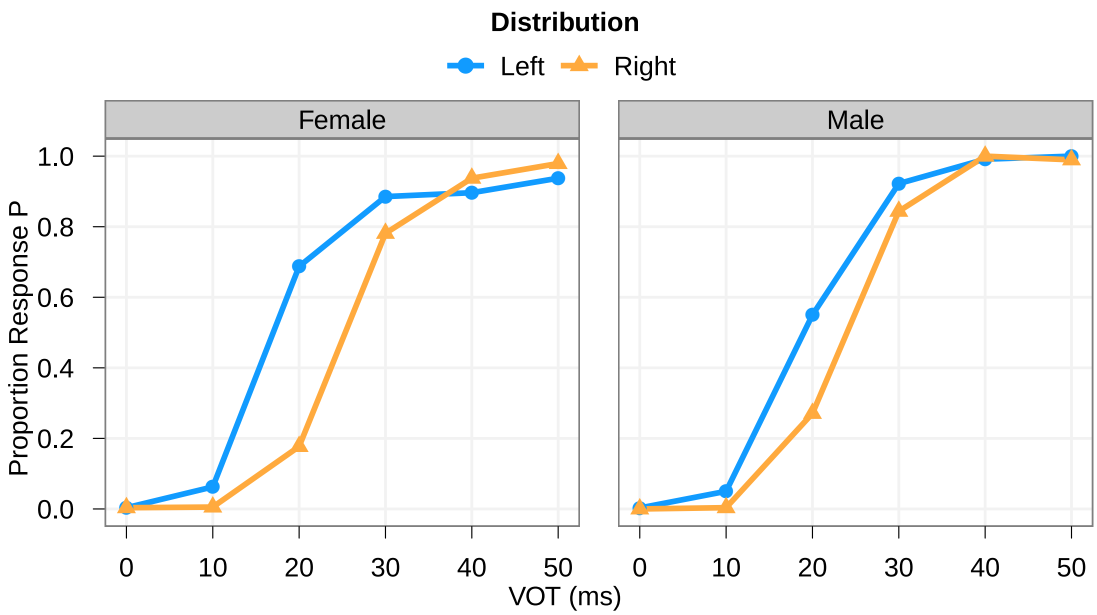

Figure 2.3: Experiment 1 distribution effect for each talker.

the right distribution. The learning analysis showed that this effect is due to learning and not *a priori* differences between the groups: the distribution effect is larger on the second day, when participants have had more exposure to the distributions. This also suggests that participants retained the boundary learning from the first day across the week between the two experiment days, which in turn suggests that the boundary shifting may have been talker-specific (listeners were certainly exposed to many other talkers and VOT exemplars in the week between the two sessions). Since this paradigm is effective in shifting category boundaries, it can be used in the remaining experiments that test the specificity and generalization of perceptual learning.

CHAPTER 3
PERCEPTUAL LEARNING FOR PHONEMES AND WORDS

3.1 Experiment 2

Experiment 2 addresses Aim 1: to test phoneme-specific perceptual learning. That is, are listeners able to learn different boundaries for a single feature contrast in multiple phonemic contexts? For example, the distinction between /b/ and /p/ relies on the same feature (voicing) as the distinction between /d/ and /t/. Listeners may be able to learn different voicing boundaries for these two contrasts (phoneme-specific learning), or they may only be able to learn a single voicing boundary that generalizes across phonemic context. Thus far the research related to this question has been inconclusive or suggested that perceptually learned boundary shifts generalize beyond the phonemic contexts in which they are learned (Kraljic & Samuel, 2006; Theodore & Miller, 2010; Clarke & Luce, 2005; Maye, Weiss, & Aslin, 2008; McQueen & Mitterer, 2005). However, these studies have not trained listeners on multiple conflicting boundaries simultaneously. It is possible that generalization is the default pattern when listeners do not have a reason to employ phoneme-specific boundary learning, but that this type of specificity is still possible in the right learning environment.

As discussed in the introduction, phoneme-specific learning would have important implications for models of speech perception. Phoneme-specific learning would indicate that abstraction at the feature level is not necessary for perceptual learning, and would be most compatible with models that do not rely on sub-lexical abstraction (such as Exemplar models, normalization, and parsing theories). Models like TRACE and MERGE would also be able to account for phoneme specific learning, but this level of specificity would mean that abstract units at a sub-phonemic level (like features) are superfluous for perceptual learning. [1] Generalization across phonemes would support a feature-like or sub-phonemic level of abstraction that could be accounted for by TRACE or MERGE but not by models lacking abstract units at a sub-lexical level.

3.1.1 Method

3.1.1.1 Design

Experiment 2 tested whether listeners can show evidence of phoneme-specific boundary shifts driven by unsupervised perceptual learning. Voicing continua with one place of articulation (e.g. bilabial) were shifted to the left and voicing continua with another place (e.g. coronal) were shifted to the right. The shift direction for each place of articulation was counterbalanced across participants. We used three item-sets with two continua each (one at each place). Each item-set consisted of a four-way contrast between the endpoints of coronal and bilabial continua (e.g. *beer*/*pier* and *deer*/*tear*. The words in each item-set are shown in Table 3.1. This four-way contrast design was used so that participants would have a reason to pay attention to place of articulation as well as VOT, which we thought might facilitate learning of the different voicing contrasts at each place. If the words were *beer*/*pier* and *dart*/*tart*, participants could ignore place of articulation and still discriminate the words. In Appendix A we report a previous experiment that did not use four-way contrast design and did not work as well (Experiment 2A).

For the distribution shift manipulation, the left and right distributions were centered at the same steps used in Experiment 1, but the number of repetitions at each step was modified to accommodate the different number of continua used. In addition, unlike the distributions in Experiment 1, both of the new distributions extended the full width of the VOT range used in the experiment, so there were no steps with 0 repetitions

[1] While TRACE and MERGE do not have such units, as abstractionist models they share this spirit.

Table 3.1: Experiment 2 stimulus items.

/b/	/p/	/d/	/t/
bart	part	dart	tart
beer	pier	deer	tear
bot	pot	dot	tot

Table 3.2: Experiment 2 VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	1	12	24	12	2	2	12	24	12	1	1	1
Right-Shifted Distribution	1	1	1	12	24	12	2	2	12	24	12	1

in either distribution. This was done to eliminate the possibility that boundary shifts could occur because of range differences between the distributions. Here there are the same number of repetitions at the endpoints of both distributions, so any boundary shifts should be due to the difference in location of the prototype steps instead of any differences at the endpoints. Finally, the distributions were applied to each individual continuum rather than across continua, which was possible because we did not have as many continua as in Experiment 1. The numbers of exemplars presented at each VOT step each continuum in the two distributions is shown in Table 3.2.

Participants completed a total of 624 trials per day, or 312 trials in each distributions/place of articulation. These were split evenly between the three continua at each place, so there were 104 trials per continuum in each distribution. The experiment was run over two sessions held approximately a week apart (1248 trials all together). Listeners heard the continua shifted in the same direction on the second day as they had on the first.

3.1.1.2 Participants

Participants were 22 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay. All participants were monolingual native English speakers who reported both normal hearing and vision that was either normal or corrected-to-normal. Informed consent was obtained in accordance with University and APA standards. 20 participants completed both sessions of the study, and 2 participants completed only one session. These 2 participants were excluded from analysis.

3.1.1.3 Stimuli

3.1.1.3.1 Auditory Stimuli

Auditory stimuli consisted of six twelve-step VOT continua ranging from -30 to 80ms. We used three /b/ to /p/ continua and three /d/ to /t/ continua. The continua were created in the same manner as those in Experiment 1, by cross-splicing recordings of natural speech. The recording and cross-splicing methods used to create the stimuli are described in Chapter 2. The male speaker who was recorded for Experiment 1 was also recorded for the new continua. When piloting the stimuli we discovered that the categorization boundaries for one of the continua, *deer/tear*, was located farther to the voiceless end of the continuum than the rest of the boundaries. Since our experimental manipulation relied on shifting this boundary, we wanted it to lie near the center of the continuum. In order to move the boundary towards the voiced side of the continuum we shortened the vowel in the voiced base stimulus (*deer*) by removing pitch pulses from the vowel center. VOT measurements of the completed stimuli are shown in Table 3.3.

The stimuli were piloted by six lab members using the same procedure that we used to pilot the stimuli for Experiment 1. On each pilot trial participants used a key-

Table 3.3: Experiment 2 VOT measurements.

bart/part	beer/pier	bot/pot	dart/tart	deer/tear	dot/tot
-30	-30	-31	-30	-30	-31
-22	-22	-23	-17	-20	-24
-15	-15	-15	-9	-12	-15
0	0	0	0	0	0
12	12	15	10	10	10
21	21	21	20	20	21
30	30	30	30	30	31
40	40	40	40	40	40
51	51	51	50	50	51
61	61	61	60	60	61
71	71	70	70	70	70
81	81	80	80	80	80

press to identify the given stimulus as beginning with /b/ or /p/ for the bilabial continua, or /d/ or /t/ for the coronal continua. The two types of continua were piloted in separate blocks. Each of the stimulus items was repeated three times. At the voiced endpoint, listeners correctly identified the stimulus as a /b/ on 100% of the trials and as /d/ on 97.8% of the trials. At the voiceless endpoint they identified the stimulus as a /p/ or /t/ on 100% of the trials. The point at which the identification functions crossed 50% was between steps 6 and 7 for all of the bilabial continua, and between 6 and 8 for the coronal continua. [2]

3.1.1.3.2 Visual Stimuli

Pictures representing each item listed in Table 3.1 were constructed using the same picture norming technique described for Experiment 1. Images that were included in the previous study were re-used here. The final images were approved by the author and thesis supervisor.

3.1.1.3.3 Procedure

The procedure was identical to that used in Experiment 1. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with the four images from an item-set (e.g. *beer, pier, deer,* and *tier*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial.

---

[2] Despite the vowel-shortening procedure, the *deer/tear* continua crossed the 50% point between steps 7 and 8. The other coronal continua crossed 50% between steps 6 and 7.

The Eye-Link II recorded participants' eye-movements throughout the experiment, but we did not analyze the eye-movement data because the effects were large enough to see with mouse-click data.

3.1.2 Results

The structure of the results section mirrors that used for Experiment 1 in Chapter 2, where each section addresses a different question. The same mixed-effects modeling strategy is also employed. First we address overall task performance. Then we assess perceptual learning of the distributions over the course of the experiment, our primary interest in this study. A boundary shift is predicted such that categorization data for continua trained on the left distribution (either bilabial or coronal onset) will have a boundary shifted towards the left, while data for the other continua (those trained on the right distribution) will show a boundary shifted towards the right. The final section of the results collapses across time to examine whether the distribution effect varied depending on shift-condition (which place was shifted to the left), and then whether it varied by item-set.

3.1.2.1 Task Performance

Mouse-click responses were first examined to assess overall performance on the task. Participants clicked on non-target continua items (e.g. *dart* or *tart* when the target word was *bart* or *part*) on only .38% of the trials, and all participants selected an image from the target continuum on over 97% of the trials, indicating that they were paying attention to both the auditory and visual stimuli. Performance on continua endpoints was also very good. On the voiced side (VOT steps -30 to -10ms), participants selected the /b/ image for 99.6% of the bilabial trials and the /d/ image for 99.9% of the coronal trials. On the voiceless side (VOT steps 60 to 80ms), participants selected the /p/ image for 99.9% of the bilabial trials and the /t/ image for 99.7% of the coronal trials.

3.1.2.2 Perceptual Learning

The critical analysis concerns perceptual learning. That is, were participants able to learn (over the course of the experiment) different boundaries for the same feature contrast (voicing) in two different phonemic contexts (/b/ and /p/ vs. /d/ and /t/)? To assess the change in the voicing boundaries over time we examined the effect of distribution (a within-participants variable) in each quarter of the experiment (by dividing each of the two days in half). Response (voiced or voiceless, dummy-coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution, day, and half were fixed factors, and were held constant in the different versions of models that we compared. All of these except for VOT step were coded as dummy variables with values of 0 or 1 and then centered. VOT step was coded from -30 to 80 (in 10ms increments) and centered. Participant and continuum were treated as random-effects. One version of the model had only random intercepts for participants and another had random intercepts for both participants and continua. A third variant with random slopes for participants and random intercepts for words failed to converge. Including continua as a random effect in addition to participants improved the fit of the model ($\chi^2(2)$=712.25, p<.0001), so this model is the one we report.

The full model for distribution learning over time is reported in Table 3.4. There was a main effect of distribution (β=-1.21, p<.0001) which indicates that there was a boundary difference between the two distribution groups. Participants in the left-shifted condition showed a boundary to the left of those in the right-shifted condition, as can be seen in Figure 3.1. The critical interactions were between distribution and day (β=-.55, p<.005) and distribution and half (β=-.64, p<.001), both of which suggest that the boundary shift was different on the first and second day of the experiment and during the first and second half of each day. This could be indicative that some perceptual learning occurred, as one would expect the effect of distribution to grow over the course

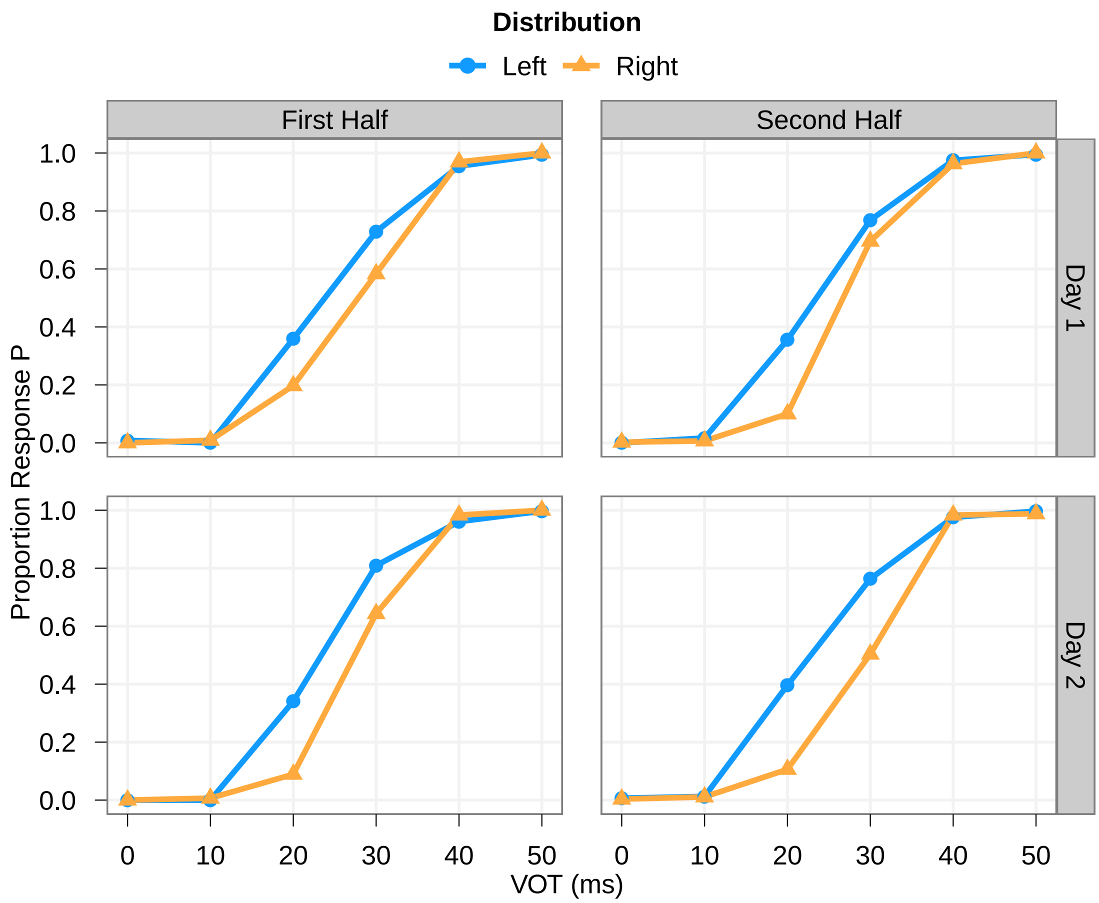

Figure 3.1: Experiment 2 distribution effect by experiment quarter.

of the experiment (as can be observed across panels of Figure 3.1).

Since the main analysis revealed interactions between distribution, day, and half, our next step was to run simple effects analyses to examine these in more detail. We particularly wanted to chart when (across days and halves) an effect of distribution could be observed in order to determine how rapidly participants learned the distributions.

To run the simple effects analysis on day and half we split the data by day and applied the previous model (without day as factor) to each subset of the data. These two models are reported in Table 3.5. On the first day there was a main effect of distribution

Table 3.4: Experiment 2 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	−0.33	0.57	−0.6	>0.6
Distribution	−1.21	0.10	−12.3	**<.0001**
Day	−0.16	0.10	−1.6	>0.1
Half	−0.12	0.10	−1.2	>0.2
VOT Step	0.29	0.01	45.0	**<.0001**
Distribution x Day	−0.55	0.19	−2.8	**<.01**
Distribution x Half	−0.64	0.19	−3.3	**<.001**
Day x Half	0.24	0.19	1.2	>0.2
Distribution x Step	0.03	0.01	2.3	**<.05**
Day x Step	0.00	0.01	0.1	>0.9
Half x Step	−0.02	0.01	−1.7	>0.1
Distribution x Day x Half	0.46	0.39	1.2	>0.2
Distribution x Day x Step	−0.03	0.02	−1.2	>0.2
Distribution x Half x Step	−0.04	0.02	−1.8	>0.1
Day x Half x Step	−0.05	0.02	−2.2	**<.05**
Distribution x Day x Half x Step	0.00	0.04	0.0	>0.9

Note: The maximum correlation among the fixed factors was r=.16, between distribution and step.

(β=-1.00, p<.0001), but there was also an interaction between distribution and half (β=-.92, p<.002). This suggests that the boundary shift may be changing from the first to the second half of this (day 1) experiment session as participants were exposed to the shifted VOT distributions. A follow-up analysis (Table 3.6) showed that the effect of distribution was significant and in the same direction for both halves (Half 1: β=-.52, p<.01; Half 2: β=.070, p<.003), suggesting that the distributions were rapidly learned during the first half, but that learning continued during the second half as well.

In contrast, on the second day there was an effect of distribution, (β=-1.42, p<.0001) but no interaction between distribution and half (β=-.33, p>.2) While the boundary shift (distribution effect) was present on the second day, the lack of interaction between distribution and half suggests that there was no further learning taking place from the first to the second half of this day. Critically, these analyses revealed that there were boundary shifts on both days (the distribution effects), and that the boundary shift was likely due to learning (since it differed by day and by half on the first day).

These results indicate that responses were influenced by distribution even in the first quarter of the experiment, when listeners had not had much exposure to the distributions. While other experiments on perceptual learning have found that boundary shifts occur relatively quickly, this experiment suggests that this is also the case in our unsupervised perceptual learning paradigm.

3.1.2.3 Place Direction Condition

Since we counterbalanced the direction that each place of articulation was shifted across participants, this analysis asks whether this interacted with the distribution effect; that is, whether distributional learning worked equally well in both directions for both places. This analysis is similar to the prior learning analysis, but as there were similar learning effects across day and half, we simplified by collapsing across day and half. Shift-condition (which place went to the left, a between-participants variable) was

Table 3.5: Experiment 2 simple effects for days one and two.

		Coef β	SE(β)	**z**	**p**
Day 1	Intercept	−0.28	0.66	−0.4	>0.7
	Distribution	−1.00	0.14	−7.0	**<.0001**
	Half	−0.22	0.14	−1.6	>0.1
	VOT Step	0.31	0.01	31.0	**<.0001**
	Distribution x Half	−0.92	0.28	−3.2	**<.01**
	Distribution x Step	0.04	0.02	2.6	**<.01**
	Half x Step	0.00	0.02	0.3	>0.8
	Distribution x Half x Step	−0.04	0.03	−1.2	>0.2
Day 2	Intercept	−0.39	0.49	−0.8	>0.4
	Distribution	−1.42	0.14	−10.4	**<.0001**
	Half	0.00	0.13	0.0	>0.9
	VOT Step	0.28	0.01	32.5	**<.0001**
	Distribution x Half	−0.33	0.27	−1.2	>0.2
	Distribution x Step	0.01	0.02	0.8	>0.4
	Half x Step	−0.04	0.01	−2.8	**<.01**
	Distribution x Half x Step	−0.03	0.03	−1.1	>0.3

Note: The maximum correlations were r=.21 for day 1 and r=.13 for day 2, and were between distribution and step on both days.

Table 3.6: Experiment 2 simple effects for each half of day one.

		Coef β	SE(β)	z	p
First Half	Intercept	−0.18	0.70	−0.3	>0.8
	Distribution	−0.52	0.20	−2.6	**<.01**
	VOT Step	0.31	0.01	22.3	**<.0001**
	Distribution x Step	0.07	0.02	3.0	**<.01**
Second Half	Intercept	−0.40	0.65	−0.6	>0.5
	Distribution	−1.48	0.21	−6.9	**<.0001**
	VOT Step	0.31	0.01	21.4	**<.0001**
	Distribution x Step	0.02	0.02	0.8	>0.5

Note: The correlation between the fixed factors was r=.32 for the first half and r=.13 for the second half.

dummy-coded as 0 (bilabial left) or 1 (coronal left) and centered. The models with random slopes failed to converge and the model including continuum as a random factor was better than the version without this factor, (χ^2(2)=397.11, p<.0001), this is the model we report (see Table 3.7).

The shift-condition model showed a main effect of distribution (β=-1.61, p<.0001), which we saw in the previous analyses, and also an effect of shift-condition (β=.72, p<.04). This indicates that responses differed based on which onset was shifted to the left (a between-participants comparison). There was also a marginal interaction between distribution and shift-condition (β=-3.27, p<.07), which suggests that there might be differences in the degree of boundary shifting for the two different groups (coronals left and bilabials left). Figure 3.2 shows the proportion of voiceless (/p/ or /t/) responses by distribution in the two different shift-conditions. The boundary difference is in the predicted direction (consistent with the VOT distributions) for the group of participants

Table 3.7: Experiment 2 shift-condition model.

	Coef β	SE(β)	**z**	**p**
Intercept	−0.22	0.47	−0.5	>0.6
Distribution	−1.61	0.14	−11.4	**<.0001**
Shift-condition	0.72	0.34	2.1	**<.05**
VOT Step	0.29	0.01	43.2	**<.0001**
Distribution x Condition	−3.27	1.74	−1.9	>0.1
Distribution x Step	0.01	0.01	1.2	>0.2
Shift-condition x Step	0.04	0.01	3.4	**<.001**
Distribution x Condition x Step	−0.13	0.03	−5.1	**<.0001**

Note: The maximum correlation among the fixed factors was r=-.033, between distribution and shift-condition.

who heard the bilabial continua shifted to the left, but the effect appears to be either null or reversed in the group that heard the coronal shifted to the left. Because of the marginal distribution by shift-condition interaction we ran simple effects analyses to determine whether the distribution manipulation was successful for both conditions.

The full results of the models for each shift-condition group are shown in Table 3.8. For the group that heard bilabials on the left (and coronals on the right), there was the predicted effect of distribution (β=-3.05, p<.0002), with the left-distribution continua (/b/-initial) showing a left-shifted boundary relative to the right-shifted continua (/d/-initial). For the group that heard coronals on the left (and bilabials on the right) we found no effect of distribution (β=.19 , p>.8), so we have no evidence that these participants shifted their categorization boundaries according to the VOT distributions they were exposed to.

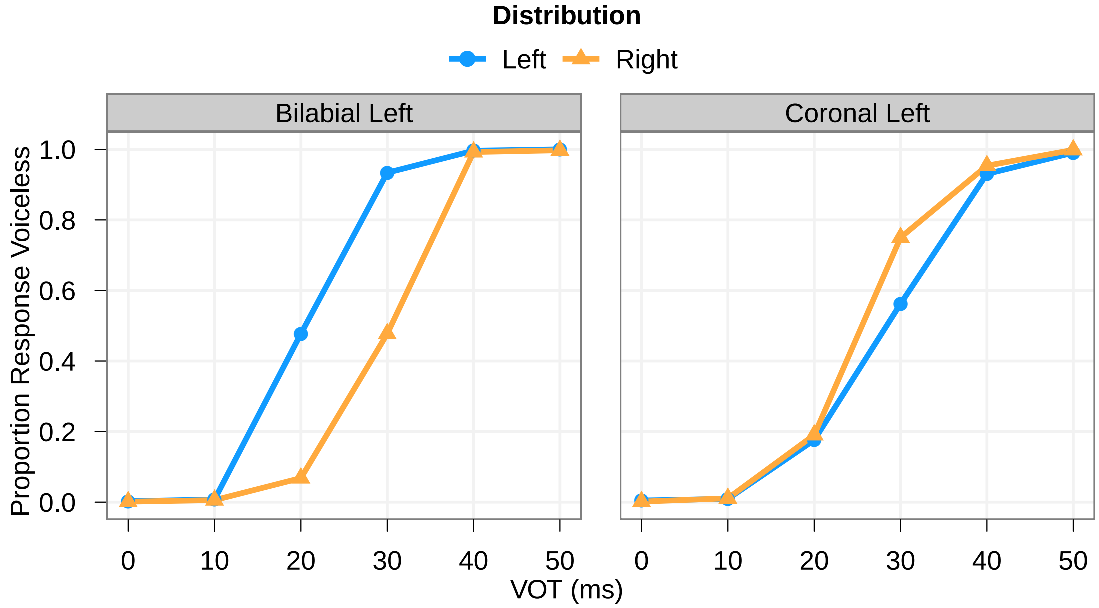

Figure 3.2: Experiment 2 distribution effect by shift-condition (which place was shifted to the left).

Table 3.8: Experiment 2 simple effects for shift-conditions.

		Coef β	SE(β)	**z**	**p**
Bilabials Left	Intercept	0.09	0.46	0.2	>0.8
	Distribution	−3.05	0.80	−3.8	**<.001**
	VOT Step	0.31	0.01	30.9	**<.0001**
	Distribution x Step	−0.04	0.02	−2.1	**<.05**
Coronals Left	Intercept	−0.62	0.51	−1.2	>0.2
	Distribution	0.19	0.90	0.2	>0.8
	VOT Step	0.27	0.01	30.4	**<.0001**
	Distribution x Step	0.08	0.02	4.8	**<.0001**

Note: The maximum correlation between the fixed factors was r=-.045 for the bilabial left model and r=.071 for the coronal left model.

3.1.3 Discussion

Experiment 2 aimed to test whether place-specific perceptual learning is possible. We did this by manipulating the VOT distributions of continua that shared the same feature contrast (voicing) but differed in place of articulation. Overall, participants showed categorization patterns consistent with the distributions they heard for each place, but this seemed to be driven by the group that heard the bilabial words shifted to the left since there was no effect of distribution for the other half of the participants.

One possible explanation for this result is that the auditory stimuli contained cues to voicing other than VOT, and these secondary cues (like F2) made it difficult to shift the coronal continua to the left and bilabial continua to the right. Other studies have found that natural-speech continua can have extra cues that interfere with the intended manipulation (Toscano & McMurray, 2011). Coronals have higher F2 values than labials and in typical, untrained performance have a voicing boundary to the right of that for labials (Sawusch & Pisoni, 1974). This could make it difficult for listeners to learn that coronals have a boundary shifted to the left of the labial boundary. This result is consistent with our piloting results as well, which indicated that the *deer*/*tear* continuum had a boundary that was already shifted towards the voiceless end of the continuum. At this time we have not measured secondary cues (such as F1 and F2) in our stimuli, but this is one way we might verify that this explanation of our results is correct. When we plotted the effect of distribution for each of the three item-sets in the two conditions, shown in Figure 3.3, we saw that the continua in one of the item-sets showed a boundary pattern consistent with the VOT distributions in both of the conditions, but the other two did not.

While the evidence in favor of phoneme or place-specific learning for boundaries is somewhat mitigated by the differences between the two conditions, the results of this experiment still suggest that participants can learn multiple boundaries for a sin-

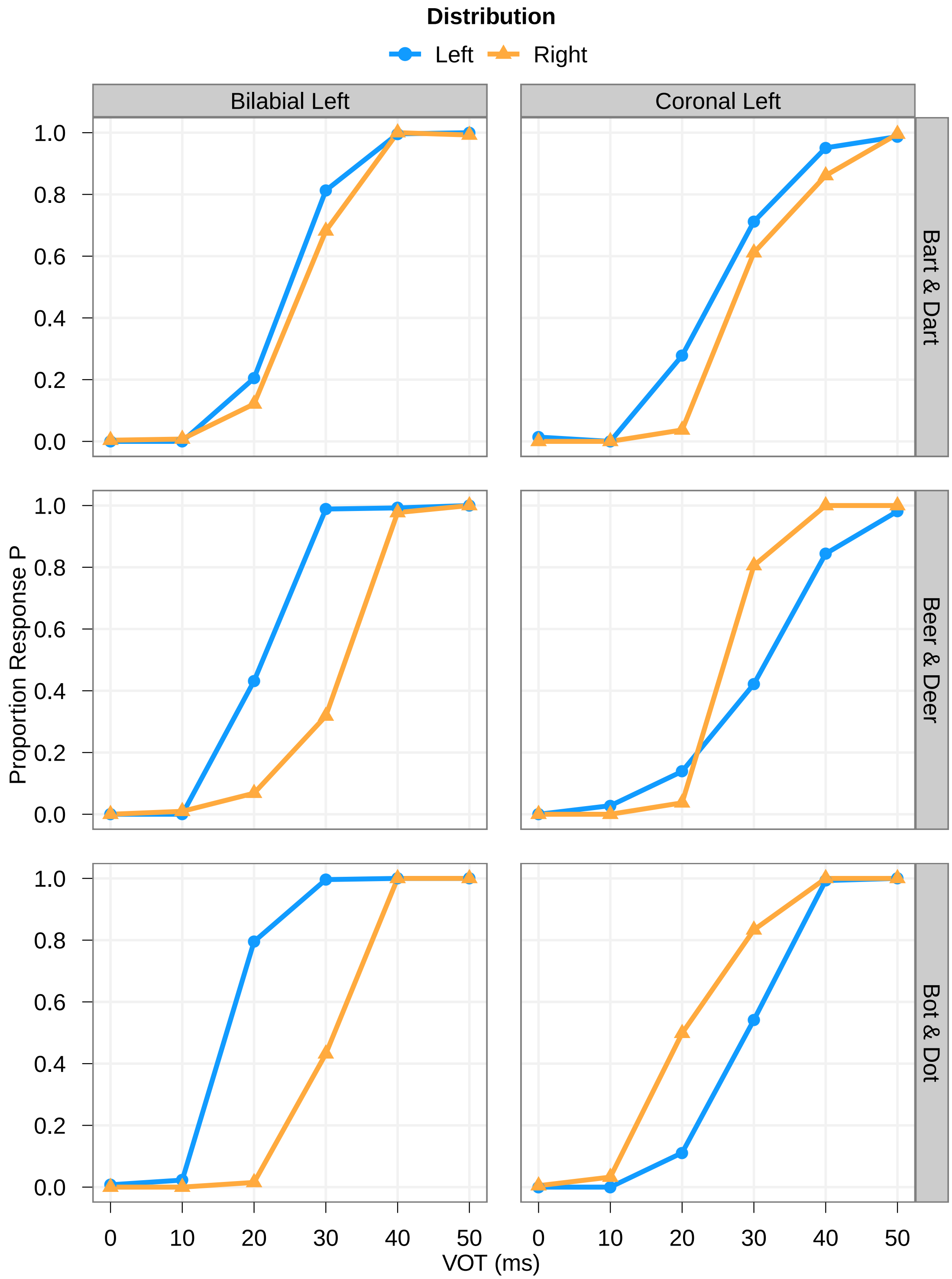

Figure 3.3: Experiment 2 distribution effect by shift-condition and item-set.

gle feature contrast, rather than a single boundary that generalizes to all contrasts relying on that feature. If participants can learn multiple boundaries, then it is possible that boundaries are even more specific than we have shown here, and could even apply to contrasts between individual lexical items. Another issue is the degree to which listeners generalize the boundaries that they learn. Previous studies have found evidence that voicing boundaries generalize across place, but perhaps boundaries are in fact lexically specific, and generalize more to words that share a place onset and less to words with different onsets. This possibility will be investigated in Experiment 3.

3.2 Experiment 3

Experiment 2 suggests that people can learn voicing boundaries specific to particular places of articulation, but some previous research has suggested that perceptually learned boundary shifts generalize to new phonemic contexts (Kraljic & Samuel, 2006; Theodore & Miller, 2010; Clarke & Luce, 2005; McQueen & Mitterer, 2005). This raises the possibility that generalization is determined by the degree of feature overlap between lexical items. Thus, Experiment 3 addresses Aim 2: to test boundary generalization to different phonemic contexts that rely on the same feature contrast. Experiment 2 indicates that despite the evidence for generalization, perceptually learned boundaries may be more context-specific than previous research has shown. If this is the case, we would expect to see significant boundary generalization to untrained words in the same phonemic context as the trained words, and little generalization (though possibly still some) to words in a different phonemic context.

This kind of learning would have implications similar to those of the previous experiment. Models like TRACE and MERGE have abstract sub-lexical representations at the phoneme level that would allow them to account for phoneme-specific learning. Additionally, while these models do not have feature-level abstractions, they could be adapted to have hierarchical abstraction (features, phonemes, and words), which would be consistent with their current architecture. Feature-like or sub-phonemic abstraction would be necessary to account for generalization across phonemes. Phoneme-specific learning, however, would indicate that abstraction at the feature level is not necessary for perceptual learning, and would also be compatible with models that do not rely on other sub-lexical abstraction (such as Exemplar models, normalization, and parsing theories).

3.2.1 Method

3.2.1.1 Design

Experiment 3 had a design similar to that of Experiments 1 and 2, but with some modifications to address the issue of generalization. Although we describe training and testing trials, we used the same type of implicit training as before, and participants received no instructions distinguishing the generalization testing trials from the training trials. As in Experiment 2, listeners heard VOT distributions that varied by place of articulation, but they were exposed to the distribution of only one place per day. Day 1 training-place (bilabial or coronal) and shift-condition (left or right) were counterbalanced across participants, and the untrained-place and distribution were trained on the second day.

The distributions used for training exposure in Experiment 3 are shown in Table 3.9. These distributions held across the continua for a given place of articulation, not within each continuum. After exposure to the training place and distribution for the day, listeners completed generalization test trials for some continua with a different place of articulation (untrained-place trials) and others with the trained-place of articulation (trained-place trials), but in different lexical items. Generalization testing was done with a flat distribution (an equal number of repetitions for each tested VOT step). Testing items were the same on both days and are listed in Table 3.10 along with both sets of training items.

Table 3.9: Experiment 3 VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	3	25	52	25	4	4	25	52	25	3	3	3
Right-Shifted Distribution	3	3	3	25	52	25	4	4	25	52	25	3

Table 3.10: Experiment 3 stimulus items.

Set Type	Voiced	Voiceless	/l/	/r/
Bilabial Training	beach	peach	lace	race
	bin	pin	lake	rake
	bug	pug	lei	ray
	bear	pear	lock	rock
Coronal Training	dot	tot	link	rink
	dune	tune	lip	rip
	dime	time	lute	root
	dent	tent	loom	room
Testing	bath	path	lane	rain
	beer	pier	list	wrist
	dart	tart	lamp	ramp
	deer	tear	leaf	reef

Note: Images for *beach/peach, beer/pier, dart/tart,* and *dot/tot* were also used in previous experiments.

Participants completed 224 experimental training trials per day, divided evenly between the four training continua for a given place, so they heard 56 experimental training trials per continuum each day. Participants also heard an equal number of filler trials which were divided evenly between item-sets and between /l/ and /r/ onset fillers. Participants completed a total of 448 training trials on each day of the two-day experiment.

Generalization testing was limited to a subset of VOT steps (-10 to 60) to increase the number of repetitions possible at each step while limiting the total number of generalization trials. All the eight VOT steps used for testing were repeated three times (a flat distribution) for each of the generalization item-sets. Two of the testing item-sets were coronal and the other two were bilabial, so there were six repetitions at each VOT step for each place. There were 96 experimental testing trials in all and an equal number of filler testing trials, evenly divided between the /l/ and /r/ filler items. Participants completed a total of 192 generalization testing trials per day.

3.2.1.2 Participants

Participants were 20 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay. All were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. All participants completed both sessions of the study, but due to an experimenter error one participant ran the experiment with the same stimuli on both days instead of hearing new stimuli on the second day. This participant was excluded from analysis.

3.2.1.3 Stimuli

3.2.1.3.1 Auditory Stimuli

Auditory stimuli consisted of six twelve-step VOT continua ranging from -30 to 80ms. There were three /b/ to /p/ continua and three /d/ to /t/ continua. These were created in the same manner as the continua in Experiment 1, by cross-splicing recordings of natural speech. The *beach/peach* and *bin/pin* were re-used from Experiment 1 and the *beer/pier, dart/tart,* and *dot/tot* continua were re-used from Experiment 2. The recording and cross-splicing methods used to create the stimuli are described in Chapter 2. The male speaker who was recorded for Experiments 1 and 2 was also recorded for the new continua. The vowel shortening procedure in which we removed pitch pulses from the center of the base stimulus vowel (previously described for the *deer/tear* continuum) was used for the *dent/tent* continuum because the boundary during piloting was shifted towards the voiceless end of the continuum. VOT measurements of the completed bilabial and coronal stimuli are shown in Tables 3.11 and 3.12, respectively.

Each continuum was piloted by four to six lab members using the same procedure that we used to pilot the stimuli for Experiments 1 and 2. On each pilot trial participants used a key-press to identify the given stimulus as beginning with /b/ or /p/ for the bilabial continua, or /d/ or /t/ for the coronal continua. The two types of continua were piloted in separate blocks. Each of the stimulus items was repeated three or four times (with the same number of repetitions for each step of a given continuum). For the bilabial continua, listeners correctly identified the stimulus as a /b/ on 100% of the voiced endpoint trials and as /p/ on 98.9% of the voiceless endpoint trials. For the coronal continua they identified the stimulus as a /d/ or /t/ on 98.9% of the trials at each endpoint. The point at which the identification functions crossed 50% was between steps 6 and 7 for all of the continua except *deer/tear* and *dent/tent*, which crossed the 50% point between steps 7 and 8 despite the vowel shortening procedure used to shift the boundary

Table 3.11: Experiment 3 VOT measurements for bilabial onset continua.

beach/peach	bin/pin	bug/pug	bear/pear	bath/path	beer/pier
-32	-29	-32	-31	-33	-30
-23	-19	-24	-22	-24	-22
-12	-10	-8	-13	-15	-15
0	0	0	0	0	0
10	9	9	7	9	12
21	20	19	18	21	21
31	30	29	29	31	30
41	40	39	40	40	40
51	50	50	50	51	51
61	60	60	60	61	61
71	71	70	70	71	71
80	80	80	80	81	81

Table 3.12: Experiment 3 VOT measurements for coronal onset continua.

dot/tot	dune/tune	dime/time	dent/tent	dart/tart	deer/tear
-31	-30	-33	-32	-30	-30
-24	-23	-23	-24	-17	-20
-15	-14	-9	-15	-9	-12
0	0	0	0	0	0
10	10	10	10	10	10
21	20	20	20	20	20
31	30	30	30	30	30
40	39	40	40	40	40
51	50	50	50	50	50
61	60	60	60	60	60
70	69	70	70	70	70
80	80	80	80	80	80

towards the voiced endpoint.

3.2.1.3.2 Visual Stimuli

Pictures representing each item listed in Table 3.10 were constructed using the same picture norming technique described for Experiment 1. Images that were included in previous studies were re-used here. The final images were approved by the author and thesis supervisor.

3.2.1.3.3 Procedure

The procedure was identical to that used in Experiments 1 and 2. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with the four images from an item-set (e.g. *bear, pear, lock,* and *rock*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial. Training and testing trials were identical except that listeners heard different continua during the testing portion of the experiment.

Participants' eye-movements were recorded during the experiment, but since the mouse-click results showed boundary shift effects we did not analyze the eye-movement data.

3.2.2 Results

The structure of the results section mirrors that used for the previous experiments, where we used mixed-effects modeling to address a different question in each section. First we assess overall task performance. Then we examine perceptual learning of the training distributions. A boundary shift is predicted such that categorization data for continua trained on the left distribution (either bilabial or coronal onset) will have a boundary shifted towards the left, while data for the other continua (those trained on the right distribution) will show a boundary shifted towards the right. In this section we address both training day (first or second) and onset place (bilabial or coronal). Finally we examine generalization trials. For each day of training, we were interested in knowing whether the generalization testing trials would show the same boundary as the training trials, and whether the two types of generalization trials (trained and untrained-place) would share the same boundary. Because generalization on day 2 was not directly comparable to generalization on day 1 (since by the end of the second day listeners had been exposed to both distributions), we examined each day separately. For each day we examined both types of generalization trials: those that shared on onset with the training trials for that day (called the trained-place trials), and those with a different onset (called the untrained-place trials). For example, test trials from the *beer/pier* continuum would be trained-place trials for a listener trained on bilabial continua, and trials from the *deer/tear* continuum would be untrained-place trials since this continuum has a coronal onset. We asked whether each type of test trial differed from the training trials, and whether the two test types differed from each other.

3.2.2.1 Task Performance

Mouse-click responses to filler images during experimental trials (e.g. clicks on *lock* or *rock* when the stimulus was *bear* or *pear*) were examined to assess task performance. Participants selected filler images for only .12% of the experimental trials, which indicates that they were paying attention to the task and not selecting images randomly. Performance at the continuum endpoints was also very good. On the voiced side (VOT steps -30 to -10ms), participants selected the /b/ image for 99.8% of the bilabial trials and the /d/ image for 99.9% of the coronal trials. On the voiceless side (VOT steps 60 to 80ms) participants selected the /p/ image for 100% of the bilabial trials and the /t/ image for 99.1% of the coronal trials. This provides further evidence that participants were attending to both the auditory and visual stimuli.

3.2.2.2 Perceptual Learning for Training Distributions

Before assessing generalization to the different types of testing trials, we needed to establish that training was successful. To do this we first examined the effect of distribution for the training trials on both days of the experiment. Response (voiced or voiceless, dummy-coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution (left or right), and day (first or second) were fixed factors held constant in the different versions of models that we compared. Distribution and day were coded as dummy variables with values of 0 (left distribution and first day) or 1 and then centered. Because participants were trained on the two distributions on different days, distribution was a within-participants variable across the two days but a between-participants variable when each day is examined in isolation. VOT step was coded as a continuous variable ranging from -30 to 80 (in 10ms increments). This was centered as well. Participant and continuum were treated as random-effects and we compared models with the same random-effects structures used in previous analyses. The random slopes models failed to converge and including continuum as a random effect improved the model (χ^2(2)=110.77, p<.0001), so this is what we report.

The full model of Experiment 3 distribution learning over time is reported in Table 3.13. The main effect of distribution (β=-1.50, p<.0001) was of primary interest since it indicates that there was a boundary difference between the left- and right-shifted continua, such that the boundary for continua heard in the left distribution is shifted to the left relative to those heard in the right distribution. Figure 3.4 shows this effect. Because listeners were trained on a different distribution on each day, the listeners in the left dis-

Table 3.13: Experiment 3 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.25	0.35	0.7	>0.5
Distribution	−1.50	0.16	−9.1	**<.0001**
Day	0.37	0.16	2.3	**<.05**
VOT Step	0.26	0.01	28.0	**<.0001**
Distribution x Day	−0.06	0.57	−0.1	>0.9
Distribution x Step	−0.01	0.02	−0.3	>0.8
Day x Step	0.02	0.02	1.3	>0.2
Distribution x Day x Step	−0.02	0.03	−0.7	>0.5

Note: The maximum correlation among the fixed factors was r=.19, between distribution and step.

tribution group on day 1 are in the right distribution group on day 2. The perceptual learning model also showed there was an effect of day (β=.37, p<.03), which suggests that responses were influenced by experiment session. However, we saw no distribution by day interaction (β=-.06, p>.9), which would have suggested that the boundary effect differed between sessions.

Next we examined whether onset-condition (bilabial or coronal on the first day, and the opposite on the second day) affected performance on the training trials. Like distribution in this experiment, onset-condition was a within-participant variable across days but between-participants on each day. Since the previous analysis found no difference in boundary learning between the two days, we collapsed across day for this analysis. The other factors in the model were the same with the addition of day 1 training onset condition as a fixed factor. This was dummy-coded (0 for bilabial, 1 for coronal) and centered. The random slopes models failed to converge and the model includ-

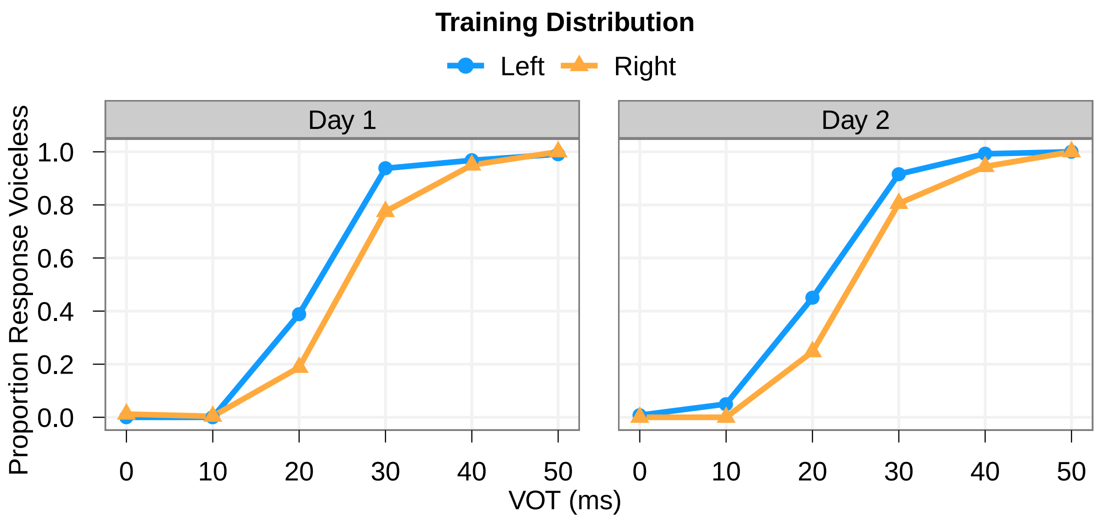

Figure 3.4: Experiment 3 training distribution effect by day. Listeners were trained on different distributions each day, so the left group on day 1 is the right group on day 2.

ing the random effect of continuum was better than the next best model ($\chi^2(2)$=112.53, p<.0001), so this is the model we report.

The full onset-condition training model is reported in Table 3.14. The model showed a main effect of distribution (β=-1.55, p<.0001) as seen in the perceptual learning model. There was no main effect of onset-condition ((β=.01, p>.9), but there was a significant interaction between distribution and onset (β=1.05, p<.003). This indicates that the boundary shift was larger for the group of participants who heard the bilabial continua on the first day, which can be seen in Figure 3.5. While both onset-condition training groups appear to show distribution effects, the effect is smaller for listeners who were initially trained on the coronals. We next ran simple effects analyses to verify that the distribution effect was significant for both groups of participants. These analyses, reported in Table 3.15, showed effects of distribution for both onset-conditions (Bilabial: β=-2.06, p<.0001; Coronal: (β=-1.09, p<.0001). Critically, this indicated a significant boundary difference for both groups of participants, meaning that the distribution

Table 3.14: Experiment 3 onset-condition model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.26	0.36	0.7	>0.5
Distribution	−1.55	0.17	−8.9	**<.0001**
Day 1 Onset	0.01	0.29	0.0	>0.9
VOT Step	0.27	0.01	24.5	**<.0001**
Distribution x Onset	1.05	0.34	3.1	**<.01**
Distribution x Step	−0.03	0.02	−1.6	>0.1
Onset x Step	−0.06	0.02	−3.0	**<.01**
Distribution x Onset x Step	0.15	0.04	3.4	**<.001**

Note: The maximum correlation among the fixed factors was r=.17, between distribution and step.

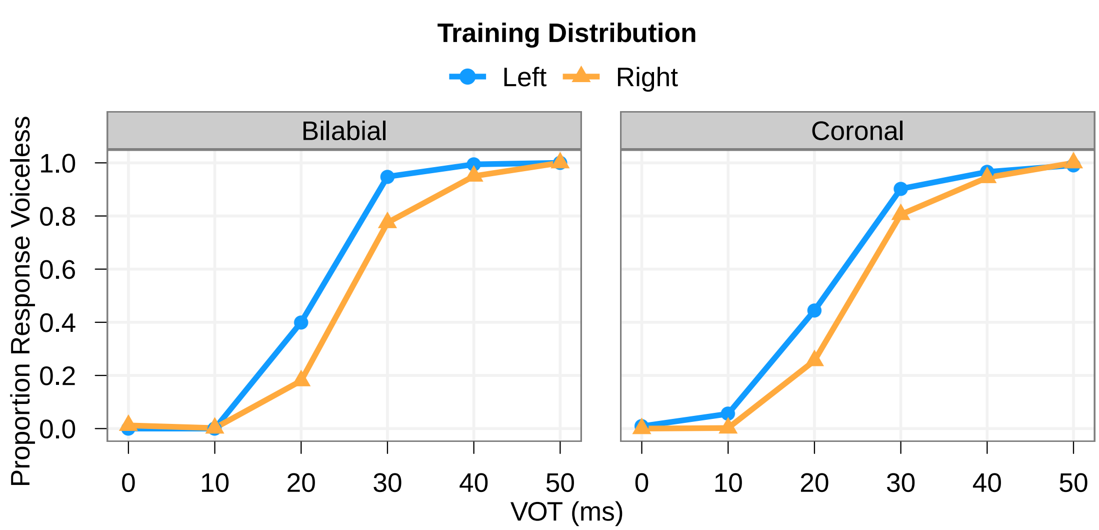

Figure 3.5: Experiment 3 distribution effect by day 1 training onset-condition.

Table 3.15: Experiment 3 simple effects for onset groups.

		Coef β	SE(β)	**z**	**p**
Bilabial Training on Day 1	Intercept	0.19	0.45	0.4	>0.7
	Distribution	−2.06	0.27	−7.6	**<.0001**
	VOT Step	0.32	0.02	15.8	**<.0001**
	Distribution x Step	−0.12	0.04	−3.1	**<.01**
Coronal Training on Day 1	Intercept	0.19	0.45	0.4	**<.0001**
	Distribution	−2.06	0.27	−7.6	**<.0001**
	VOT Step	0.32	0.02	15.8	**<.0001**
	Distribution x Step	−0.12	0.04	−3.1	**<.0001**

Note: The correlation between fixed factors was r=.09 for Bilabial training and r=.3 for Coronal training.

shifts were learnable for participants trained on either bilabial or coronal continua on the first day.

While the order of training on the two place contrasts did not affect learning, Experiment 2 showed that the direction of each place did affect learning–only participants who heard the bilabial continua in the left distribution were able to learn the boundary shifts. We assess the affect of place direction for Experiment 3 in another model, coding place direction as 0 for bilabial left and 1 for coronal left. The other factors in the model were unchanged. The random slopes models failed to converge and the model including the random effect of continuum was better than the next best model (χ^2(2)=28.43, p<.0001), so this is the model we report.

The place-direction training model is reported in Table 3.16. The model showed a main effect of distribution (β=-1.18, p<.0001) as seen in the perceptual learning model. There was also a main effect of place direction ((β=1.16, p<.0001) and an interaction be-

Table 3.16: Experiment 3 place-direction model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.49	0.23	2.1	**<.05**
Distribution	−1.18	0.19	−6.1	**<.0001**
Place Direction	1.16	0.26	4.4	**<.0001**
VOT Step	0.29	0.01	24.1	**<.0001**
Distribution x Direction	3.78	0.85	4.5	**<.0001**
Distribution x Step	0.04	0.02	1.5	>0.1
Direction x Step	0.04	0.02	1.9	>0.1
Distribution x Direction x Step	0.26	0.05	5.5	**<.0001**

Note: The maximum correlation among the fixed factors was r=.37, between distribution and step.

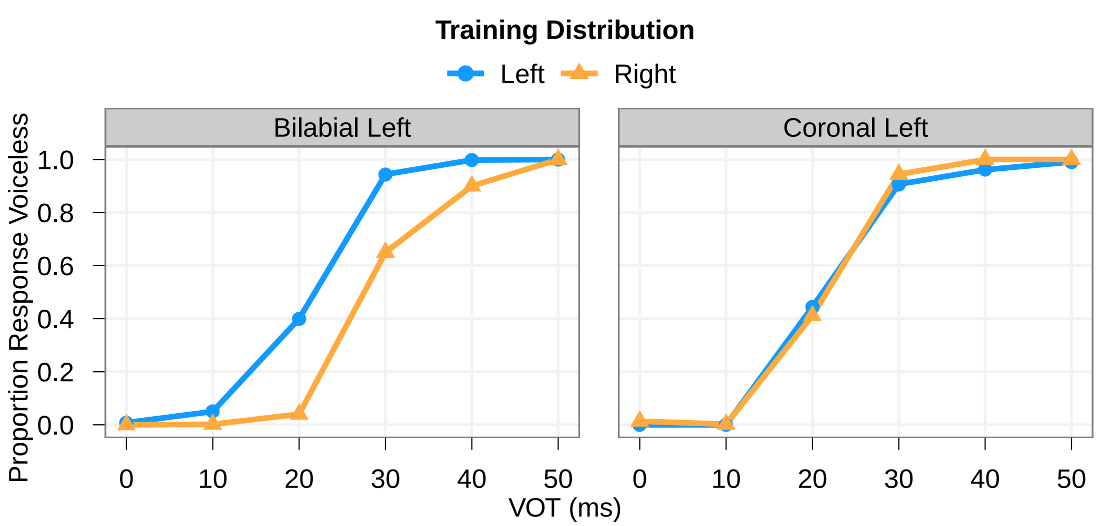

Figure 3.6: Experiment 3 distribution effect by place direction condition.

Table 3.17: Experiment 3 simple effects for place direction.

		Coef β	SE(β)	**z**	**p**
Bilabial Left	Intercept	−0.10	0.16	−0.6	>0.5
	Distribution	−2.87	0.31	−9.2	**<.0001**
	VOT Step	0.25	0.01	19.6	**<.0001**
	Distribution:Step	−0.07	0.03	−2.9	**<.01**
Coronal Left	Intercept	1.15	0.40	2.8	**<.01**
	Distribution	0.77	0.70	1.1	>0.3
	VOT Step	0.32	0.02	16.1	**<.0001**
	Distribution:Step	0.19	0.04	4.9	**<.0001**

Note: The correlation between fixed effects was r=-.06 for Bilabial left and r=.3 for Coronal left.

tween distribution and direction (β=3.78, p<.0001). These indicate that not only did responding differ between the direction conditions overall, but that the effect of distribution was also different for the two groups. This can be seen in Figure 3.6, where as in Experiment 2, the only visible boundary difference is for the listeners who heard the bilabial continua shifted left (and the coronal continua shifted right). Follow-up analyses (Table 3.17) confirmed that the distribution effect was only significant for the bilabial-left group (Bilabial left: β=-2.87, p<.0001; Coronal left: β=.77, p>.2). This suggests that blocking exposure to the two distributions does not make it easier for listeners to learn boundary shifts that conflict with secondary cues to voicing, which are inherent to place of articulation.

3.2.2.3 Generalization Trials on Day 1

Our next analysis examined generalization trials. We first assessed the trained-place trials from the first day by comparing them with the training trials for that day. Our goal here was to determine whether the boundary learned during the training phase of the first session would also be used with the generalization continua that had the same onset place as the continua listeners heard during training. This is an odd analysis because for each participant we are comparing a large number of training trials (224) with a bi-modal distribution to a small number of generalization trials (48) with a uniform distribution. Mixed-effects models are an especially appropriate approach in this case because they can handle the imbalance present in our design, accounting for the number of data points present.

As in the previous analyses, response was the dependent variable, and VOT step and training distribution were fixed factors. The new fixed factor introduced here was trial-type (training or testing), dummy coded at 0 or 1 and centered. These factors were held constant in all versions of the models we compared. A model with random slopes for participants failed to converge, and including random intercepts for continua as well as participants improved the fit of the model ($\chi^2(2)$=53.82, $p<.0001$), so this was the selected model.

The model comparing day 1 training trials and trained-place generalization trials is shown in Table 3.18. There was a main effect of training distribution (β=-1.62, $p<.0001$) but not trial-type (β=.86, $p>.2$), shown in Figure 3.7, where the training and testing trials appear to have the same boundary. Our failure to find evidence of any difference between training and testing trials suggests that listeners generalize training distributions to the trained-place generalization continua. An interaction between trial-type and step (β=.093, $p<.02$) and a marginal interaction between distribution, trial-type, and step (β=-.13, $p<.08$) suggest there were slope differences present, but we will remain fo-

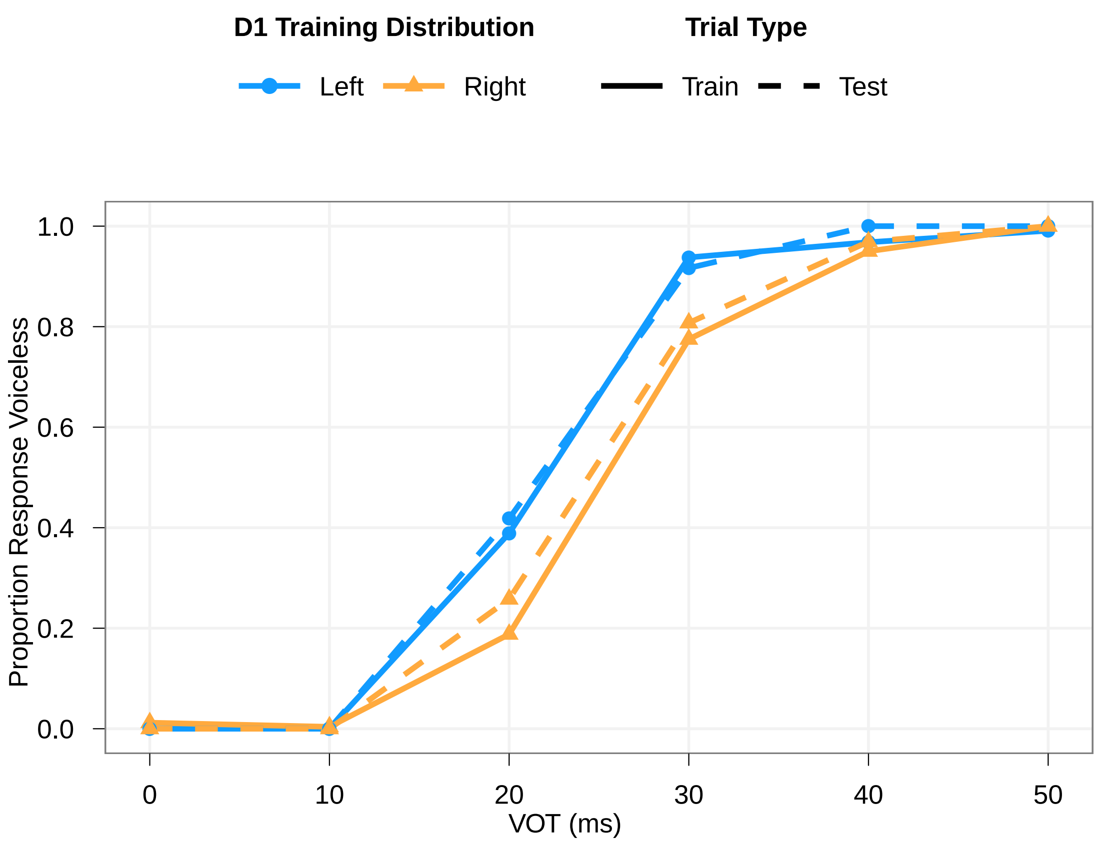

Figure 3.7: Experiment 3 day 1 training and trained-place generalization trials.

Table 3.18: Experiment 3 day 1 trained-place trials model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.30	0.34	0.9	>0.4
Training Distribution	−1.62	0.39	−4.2	**<.0001**
Trial-type	0.86	0.63	1.4	>0.2
VOT Step	0.28	0.01	22.3	**<.0001**
Distribution x Trial-type	0.18	0.51	0.4	>0.7
Distribution x Step	−0.03	0.02	−1.4	>0.2
Trial-type x Step	0.09	0.04	2.5	**<.05**
Distribution x Trial-type x Step	−0.13	0.07	−1.8	>0.1

Note: The maximum correlation among the fixed factors was r=-.037, between training distribution and trial-type.

cused on boundary differences.

While we observed no boundary differences between training trials and trained-place testing trials, we were equally interested in any differences between training and untrained-place testing trials (the generalization continua with the untrained-place of articulation). That is, we wanted to know whether boundary shift training would generalize to continua with a new place of articulation, just as it generalized to continua with the trained-place onset. The model we used to look at untrained-place generalization trials was identical to the trained-place models except that we compared training trials with untrained-place trials instead of trained-place trials.

The model comparing day 1 training and untrained-place generalization trials showed a main effect of training distribution (β=-1.39, p<.0009) but not trial-type (β=.16, p>.8), just like the trained-place model, again indicating a lack of overall difference between training and testing trials. However, there was a significant interaction between

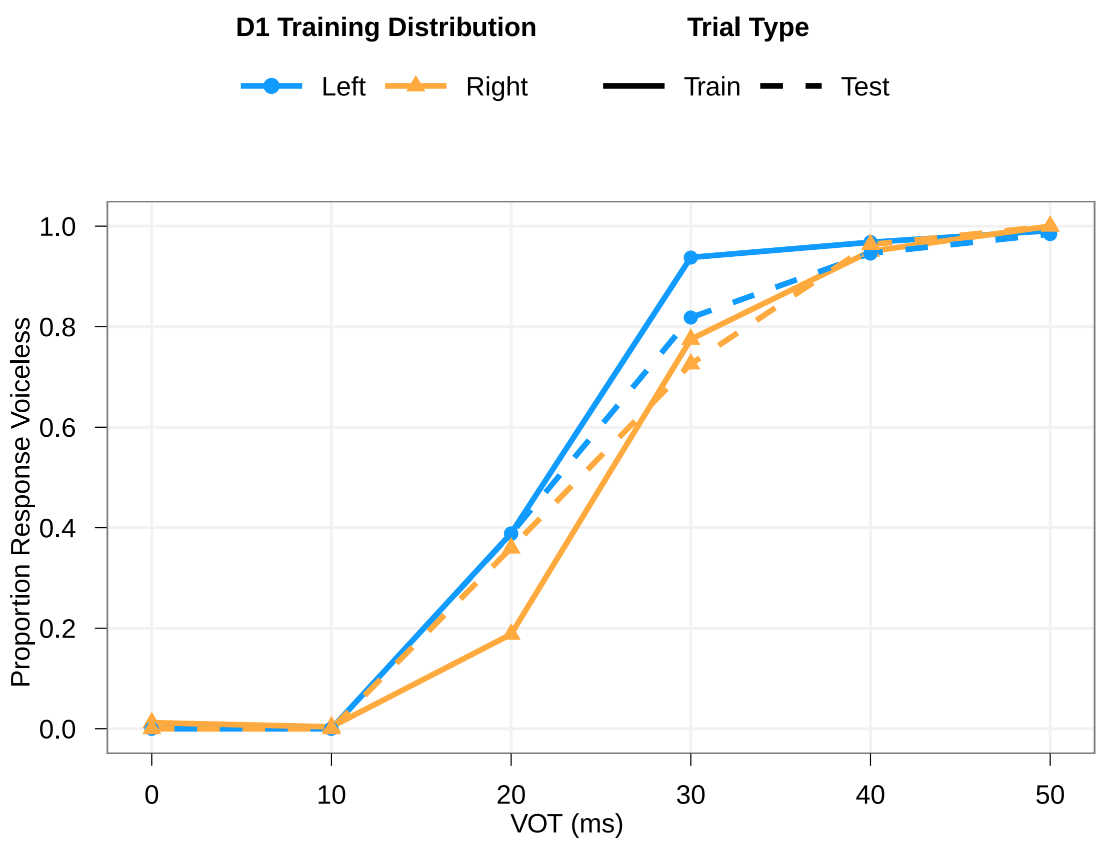

Figure 3.8: Experiment 3 day 1 training and untrained-place generalization trials.

Table 3.19: Experiment 3 day 1 untrained-place trials model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.15	0.36	0.4	>0.7
Training Distribution	−1.39	0.42	−3.3	**<.001**
Trial-type	0.16	0.64	0.3	>0.8
VOT Step	0.26	0.01	23.5	**<.0001**
Distribution x Trial-type	1.52	0.39	3.9	**<.001**
Distribution x Step	0.00	0.02	0.0	>0.9
Trial-type x Step	0.00	0.02	0.0	>0.9
Distribution x Trial-type x Step	0.03	0.04	0.8	>0.4

Note: The maximum correlation among the fixed factors was r=.057, between training distribution step.

distribution and trial-type (β=.1.51, p<.0002), which suggests that the difference between training and testing trials is not the same for the two distribution groups (the participants trained on the left distribution and those trained on the right). Figure 3.8 shows the response data from each distribution group for the two different trial-types. Although there are no obvious boundary differences, we ran simple effects analyses to follow-up on the distribution by trial-type interaction. These models, reported in Table 3.20, showed no effect of trial-type (Left: (β=-.71, p>.3; Right: (β=.87, p>.1), so the difference between training and testing trials was not significant for either distribution group.

3.2.2.4 Generalization Trials on Day 2

We assessed generalization trials for the second day in the same way that we approached generalization for the first day. Testing continua were the same on both days but because listeners were trained on new continua their second day (those with the

Table 3.20: Experiment 3 simple effects for day 1 untrained-place trials.

		Coef β	SE(β)	**z**	**p**
Left-Shift on Day 1	Intercept	0.95	0.44	2.1	**<.05**
	Trial-type	−0.71	0.71	−1.0	>0.3
	VOT Step	0.27	0.02	15.8	**<.0001**
	Trial-type x Step	−0.03	0.03	−0.9	>0.4
Right-Shift on Day 1	Intercept	−0.49	0.40	−1.2	>0.2
	Trial Type	0.87	0.64	1.3	>0.2
	VOT Step	0.26	0.02	17.2	**<.0001**
	Trial Type x Step	0.02	0.04	0.6	>0.5

Note: The correlation between the fixed factors was r=.035 for the left-shifted model and r=-.29 for the right-shifted model.

place they were not initially trained on), the untrained-place continua from the first day became the trained-place continua on the second day.

In the first analysis we compared day 2 training trials to day 2 trained-place trials with the same model structure used to examine generalization on day 1. Because distribution training condition was coded according to day 1, the distribution effect is expected to have the opposite direction from that observed for the day 1 models. The model with random slopes for participants and random intercepts for words failed to converge, and including both words and participants as random effects improved the fit of the model ($\chi^2(2)=48.21$, p<.0001), so this model is the one we report.

The day 2 trained-place trials comparison model is shown in Table 3.21. The model showed a main effect of training distribution (β=1.28, p<.007), indicating boundary differences between the two training distribution groups. As on the first day there was no effect of trial-type (β=.42, p>.5), suggesting listeners used the same boundary for

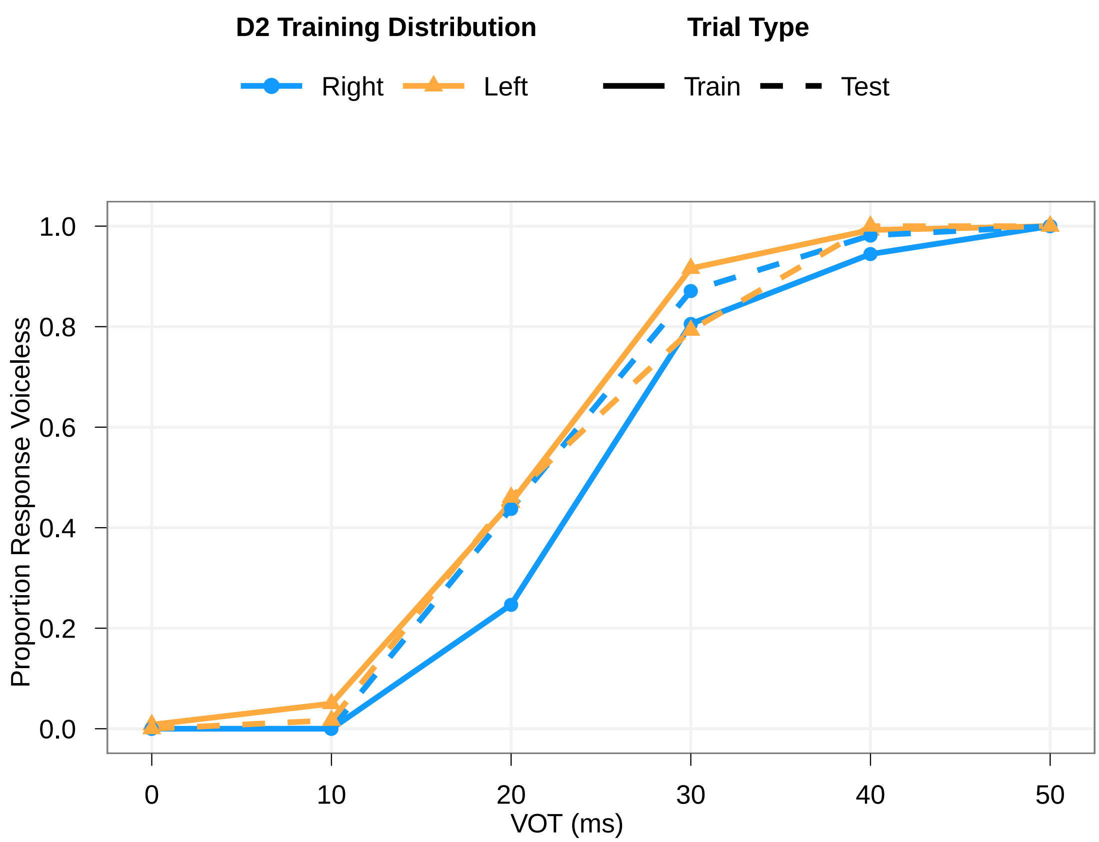

Figure 3.9: Experiment 3 day 2 training and trained-place generalization trials.

Table 3.21: Experiment 3 day 2 trained-place trials model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.49	0.37	1.3	>0.2
Day 1 Distribution	1.29	0.47	2.7	**<.01**
Trial-type	0.42	0.62	0.7	>0.5
VOT Step	0.29	0.01	21.5	**<.0001**
Distribution x Trial-type	−2.09	0.45	−4.6	**<.0001**
Distribution x Step	0.01	0.03	0.5	>0.6
Trial-type x Step	0.03	0.03	1.0	>0.3
Distribution x Trial-type x Step	−0.04	0.06	−0.8	>0.4

Note: The maximum correlation among the fixed factors was r=-.07, between training distribution and step.

the training trials and the trained-place generalization trials. There was, however, a significant interaction between distribution and trial-type (β=-2.09, p<.0001) which indicates that the differences between training and testing trials is larger for the group of listeners trained on the right distribution (on the second day) than for those trained on the left. This can be seen in Figure 3.9. Follow-up analyses (reported in Table 3.22) showed no significant effect of trial-type for either group (Left: β=-.57, p>.3; Right: β=1.51, p>.1), indicating that there was no difference between training and trained-place testing trials for either distribution group.

We compared day 2 training and untrained-place generalization trials in a similar model. The version with random slopes failed to converge, random intercepts for continua as well as participants improved the fit of the model (χ^2(2)=46.04, p<.0001). The results are reported in Table 3.23. There was an effect of distribution (β=1.03, p<.04) and a significant interaction between distribution and trial-type (β=-3.18, p<.0001). A plot of

Table 3.22: Experiment 3 simple effects for day 2 trained-place trials.

		Coef β	SE(β)	**z**	**p**
Left-Shift on Day 2	Intercept	1.09	0.38	2.9	**<.01**
	Trial-type	−0.57	0.52	−1.1	>0.3
	VOT Step	0.29	0.02	15.6	**<.0001**
	Trial-type x Step	0.02	0.04	0.4	>0.7
Right-Shift on Day 2	Intercept	−0.23	0.55	−0.4	>0.7
	Trial-type	1.51	1.04	1.5	>0.1
	VOT Step	0.31	0.02	14.9	**<.0001**
	Trial-type x Step	0.00	0.05	0.0	>0.9

Note: The correlation between the fixed factors was r=.079 for the left-shift model and r=.013 for the right-shift model.

these data, Figure 3.10, shows that that the direction of the difference between training and testing trials is reversed between the two training distribution groups. The testing trials appear shifted in the direction of the day 1 training, which suggests that listeners may have retained the day 1 distribution and are using the boundaries from day 1 for day 2 generalization untrained-place continua (which have an onset that matched the day 1 training continua). Follow-up analyses (reported in Table 3.24) showed that there was an effect of trial-type for both the left- and right-shifted groups (initially trained on the opposite distributions) (Left: β=-1.20, p<.04; Right: β=2.09, p<.03). Both distribution groups showed a significant boundary difference between training and testing trials, and the effects were in the opposite direction for the two groups. Moreover, the direction of the effects suggests that the difference may be due to the training from the first day: perhaps listeners have retained their boundaries from the training on day 1 and are using these for the day 2 untrained-place trials.

Table 3.23: Experiment 3 day 2 untrained-place trials model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.41	0.36	1.1	>0.3
Training Distribution	1.03	0.49	2.1	**<.05**
Trial-type	0.33	0.59	0.6	>0.6
VOT Step	0.29	0.01	21.6	**<.0001**
Distribution x Trial-type	−3.18	0.46	−6.9	**<.0001**
Distribution x Step	−0.01	0.03	−0.5	>0.6
Trial-type x Step	0.02	0.03	0.6	>0.6
Distribution x Trial-type x Step	−0.04	0.06	−0.7	>0.5

Note: The maximum correlation among the fixed factors was r=-.07, between training distribution and step

Table 3.24: Experiment 3 simple effects for day 2 untrained-place trials.

		Coef β	SE(β)	**z**	**p**
Left-Shift on Day 2	Intercept	0.88	0.34	2.6	**<.01**
	Trial-type	−1.20	0.57	−2.1	**<.05**
	VOT Step	0.28	0.02	15.9	**<.0001**
	Trial-type x Step	0.01	0.04	0.2	>0.9
Right-Shift on Day 2	Intercept	−0.12	0.54	−0.2	>0.8
	Trial-type	2.09	0.96	2.2	**<.05**
	VOT Step	0.32	0.02	14.8	**<.0001**
	Trial-type x Step	0.02	0.05	0.5	>0.6

Note: The correlation between the fixed factors was r=.034 for the left-shift model and r=.054 for the right-shift model.

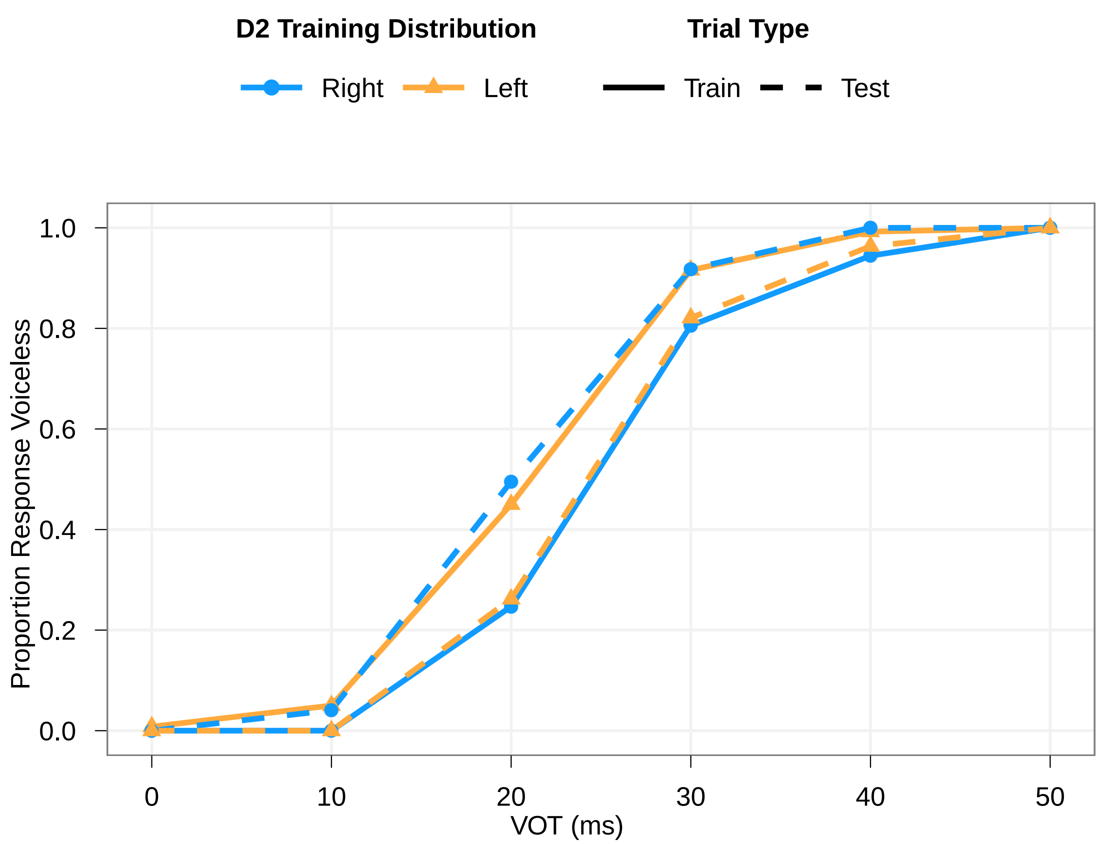

Figure 3.10: Experiment 3 day 2 training and untrained-place generalization trials. Untrained-place trials from day 2 share an onset with the training trials from day 1, which were trained in the opposite distribution.

Table 3.25: Experiment 3 day 1 training and day 2 untrained-place model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.27	0.33	0.8	>0.4
Training Distribution	−1.69	0.47	−3.6	**<.001**
Trial-type	0.73	0.53	1.4	>0.2
VOT Step	0.27	0.01	23.2	**<.0001**
Distribution x Trial-type	0.21	0.46	0.4	>0.7
Distribution x Step	−0.01	0.02	−0.6	>0.6
Trial-type x Step	0.04	0.03	1.5	>0.1
Distribution x Trial-type x Step	−0.05	0.06	−0.9	>0.3

Note: The maximum correlation among the fixed factors was r=.071 between trial-type and step.

To test this hypothesis we compared the day 2 untrained-place trials with the day 1 training trials. The modeling procedure and structure of the model was identical to that used for the previous analyses. The model with random slopes for participants and random intercepts for words failed to converge, and including both words and participants as random effects improved the fit of the model (χ^2(2)=34.59, p<.0001), so this model is the one we report.

The model comparing day 1 training and day 2 untrained-place generalization is shown in Table 3.25. There was an effect of training distribution (β=-1.69, p<.0004) but no effect of trial-type (β=.73, p>.2) or trial-type by distribution interaction (β=.21, p>.7). Thus, we have no evidence of a difference between the day 1 training data and the day 2 untrained-place testing data. A plot of these data is shown in Figure 3.11. Critically, the day 2 untrained-place continua had the same place as the day 1 training continua. The lack of a difference between the boundaries for these two types of con-

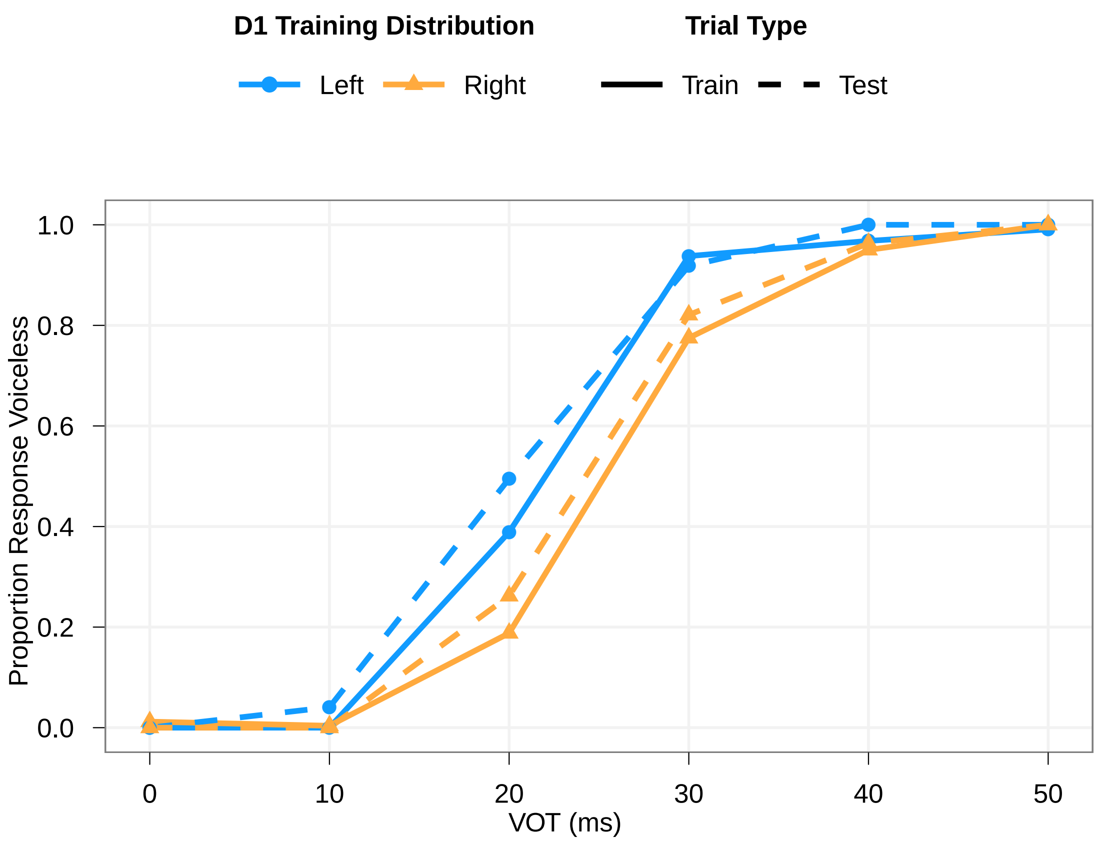

Figure 3.11: Experiment 3 day 1 training and day 2 untrained-place generalization trials. Untrained-place trials from day 2 share an onset with the training trials from day 1.

tinua, coupled with the difference seen between these test trials and the day 2 training trials, suggests that participants have retained the boundaries they learned on the first day and are using those boundaries (instead of the more recently trained boundaries) for the test continua that have the initially trained onset.

3.2.3 Discussion

Experiment 3 aimed to test boundary generalization to different phonemic contexts that rely on the same feature contrast. We did this by manipulating the VOT distributions of continua with a single onset place, and testing generalization to new continua with the same onset place and a different onset place. On the first day, participants generalized perceptually learned voicing boundaries to untrained continua with both the same and different places of articulation. This is consistent with previous research showing generalization across phonemic contexts relying on the same feature contrast (Kraljic & Samuel, 2006; Theodore & Miller, 2010; Clarke & Luce, 2005; Maye, Weiss, & Aslin, 2008; McQueen & Mitterer, 2005). However, on the second day, listeners were trained on a new boundary for continua with a different onset place than the one they were trained on the first day. Generalization testing trials for the second day showed that listeners generalized this new boundary to the testing continua with the same place onset, but not to those with a different place. Instead, listeners used the boundary from the first day for these continua.

The results of this experiment suggest that participants can generalize boundaries across different phonemic contrasts relying on the same feature contrast, but they can also learn and retain multiple boundaries for a feature contrast in different phonemic contexts. It appears that listeners generalize boundaries when they do not have a reason to do otherwise (as on the first day), but are able to use more specific boundaries when they have learned them previously. Experiment 4 investigates the degree of specificity that is possible with this type of learning, asking if listeners are able to learn different boundaries for individual word pairs that have the same onsets.

3.3 Experiment 4

Experiments 2 and 3 offer strong support for the possibility that people can learn voicing boundaries specific to particular places of articulation. This raises the possibility that learning could also be conditionalized on other sources of information, possibly even more specific. Thus, Experiment 4 addresses Aim 3: to test whether word-specific perceptual learning is possible. Exemplar and parsing theories could both account for word-specific learning since acoustic information is linked directly to lexical representations. Theories with abstract sub-lexical representations (such as the phonemes in TRACE and Merge) would find it more difficult to explain categorization boundaries that only apply to specific words. If listeners were able to learn different voicing boundaries for words with the same onset CVs, it would support direct links from acoustic representations to the lexicon.

3.3.1 Method

3.3.1.1 Design

Participants in Experiment 4 heard words with the same consonant-vowel onsets shifted in opposite directions (e.g. *beach/peach* was shifted to the left, and *beak/peak* was shifted to the right). It is important to hold the onset of the words constant in order to test lexically-specific boundary learning, because if the words had different onsets then lexically-specific learning would look the same as sub-lexical learning. For example, if listeners learned different boundaries for *beach/peach* and *bin/pin*, then listeners might simply be learning boundaries conditionalized on the vowel context, or that were biphone specific (/bi/ and /bɪ/) instead of lexically-specific. By keeping the CV onsets the same, we eliminate these possibilities, increasing our confidence that any observed boundary shifts are lexically-specific.

For each participant, half of the continua were heard with VOTs from the left distribution (the left-shift words) and the other half from the right distribution (the right-shift words). Table 3.26 shows the words used for the experiment, which included only experimental words and no filler items. Fillers were eliminated from Experiment 4 so that items from two experimental continua sharing the same CV onset could appear together on each trial. This was done so that participants would have an incentive to track which distribution each continuum belonged with, in order to make faster decisions about which word they heard on each trial. If pairs of experimental items were instead paired with filler items, as in Experiment 1, there would be no advantage to tracking the distribution of each word pair. A previous experiment with filler items instead of item-sets with pairs of experimental items is reported in Appendix B (Experiment 4A). The items listed in Table 3.26 are divided into three sets with two minimal pairs each. For each participant, one word pair from each of these sets was randomly assigned to the left-distribution. The other pair with the same CV onset was then assigned to the right-distribution.

Table 3.27 shows the VOT distributions that were used for each continuum. In this experiment the distributions held within continuum as well as across all the continua in each distribution. The overall distributions that participants heard on each day had three times the number of repetitions shown in Table 3.27 because there were three continua in each distribution. Both of the distributions extended the full width of the VOT range used in the experiment (as in Experiment 2).

Listeners completed a total of 648 critical trials per day, 324 in each distribution (108 per continuum). Each listener participated in two experiment sessions for a total of 1296 trials per listener. The assignment of continua to distribution condition (for each participant) was kept the same for the second session, so trial order was the only difference between the two sessions.

Table 3.26: Experiment 4 stimulus items.

	left distribution		right distribution	
Participant	/b/	/p/	/b/	/p/
1	beach	peach	beak	peak
	bill	pill	bin	pin
	buck	puck	bug	pug
2	beak	peak	beach	peach
	bill	pill	bin	pin
	bug	pug	buck	puck
3	beach	peach	beak	peak
	bin	pin	bill	pill
	bug	pug	buck	pick

Note: For each participant, one continuum from each of the three CV onsets pair was randomly assigned to each of the two distributions, so this is a sample of possible distribution assignments.

Table 3.27: Experiment 4 VOT distributions

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	2	12	24	12	2	2	12	24	12	2	2	2
Right-Shifted Distribution	2	2	2	12	24	12	2	2	12	24	12	2

3.3.1.2 Participants

Participants were 22 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay. All participants were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. 20 participants completed both sessions of the study, and 2 participants completed only one session. These 2 participants were excluded from analysis.

3.3.1.3 Stimuli

3.3.1.3.1 Auditory Stimuli

Auditory stimuli consisted of six twelve-step /b/ to /p/ VOT continua ranging from -30 to 80ms. Four of the continua were originally constructed for Experiment 1 and the remaining two continua (*buck/puck* and *bug/pug*) were new. The new continua were created in the same manner as those in Experiment 1, by cross-splicing recordings of natural speech. The recording and cross-splicing methods used to create the stimuli are described in Chapter 2. The same male speaker who was recorded for Experiment 1 was also recorded for the two new continua. The recordings were made in the same location and we tried to match the recording levels as closely as possible. VOT measurements of all the stimuli used in this experiment are shown in Table 3.28.

The new stimuli were piloted using the same categorization task that was used to pilot the original continua. The *buck/puck* continuum was piloted by six lab members and the *bug/pug* continuum was piloted by seven lab members. On each pilot trial participants used a key-press to identify the given stimulus as beginning with /b/ or /p/. Each of the 24 stimulus items was repeated three or four times (consistent for each continuum and participant). At the /b/ endpoint, listeners correctly identified the stimulus as a /b/ on 100% of the trials. At the /p/ endpoint they identified the stimulus as a /p/ on 100% of the *buck/puck* trials and 95.8% of the *bug/pug* trials. The point at which

Table 3.28: Experiment 4 VOT measurements.

beach/peach	beak/peak	bill/pill	bin/pin	buck/puck	bug/pug
-32	-30	-33	-29	-31	-32
-23	-17	-25	-19	-22	-24
-12	-8	-8	-10	-11	-8
0	0	0	0	0	0
10	10	10	9	9	9
21	20	21	20	19	19
31	33	31	30	30	29
41	40	41	40	40	39
51	52	51	50	50	50
61	60	61	60	60	60
71	71	71	71	70	70
80	80	81	80	80	80

Note: Buck/puck and bug/pug were the new continua created for this experiment.

the identification functions crossed 50% was between steps 6 and 7 for both continua, which was the same crossover point seen when piloting the four other continua.

3.3.1.3.2 Visual Stimuli

Pictures representing each item listed in Table 3.26 were constructed using the same picture norming technique described for Experiment 1 in Chapter 2. Images that were included in the previous study were re-used here. The final images were approved by the author and thesis supervisor.

3.3.1.4 Procedure

The procedure was identical to that used in Experiment 1, described in Chapter 2. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with four images from two different continua sharing the same CV-onset (e.g. *beach, peach, beak,* and *peak*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial. The Eye-Link II recorded participants' eye-movements throughout the experiment, but because effects were observable in the mouse-click data we did not analyze the eye-movement data.

3.3.2 Results

The structure of the results section mirrors that used for previous experiments, where each section addresses a different question. The same mixed-effects modeling strategy is also employed. The first part of the results section addresses overall task performance. The second section assesses perceptual learning of the distributions over the course of the experiment. If listeners can lexically-specific boundaries, we predict that continua trained in the left distribution will have a boundary shifted towards the left, while the other continua (those trained on the right distribution) will have a boundary shifted towards the right. If this boundary shift is learned over the course of the experiment, this trend should increase over time. The final section of the results collapses across time in order to investigate whether the effect of distribution varied by vowel, as the six continua used had three different vowels.

Table 3.29: Experiment 4 percent correct at endpoints.

Continuum	/b/	/p/
beach/peach	99.2	100
beak/peak	99.3	93
bill/pill	98.8	100
bin/pin	99.3	100
buck/puck	99.7	100
bug/pug	98.1	100

3.3.2.1 Task Performance

Mouse-click responses were first examined to assess overall performance on the task. Participants clicked on an image from the incorrect continuum (e.g. *beach* or *peach* when the stimulus was *beak* or *peak*) on an average of .36% of the trials, and the lowest individual accuracy score was 98.9%. Performance at continuum endpoints was also excellent: /b/ steps (-30 to -10ms) averaged 99.5% correct, and /p/ steps (60 to 80ms) averaged 99.2%. Performance was high for all continua (reported in Table 3.29).

3.3.2.2 Perceptual Learning

The most important analysis concerns perceptual learning, asking whether lexically specific boundary shifts emerged over the course of the experiment. To look at the learning effect over time we looked at the effect of distribution, a within-participant variable, day, and half. This model is the same as the perceptual learning model constructed for Experiment 2. Response (/b/ or /p/, coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution, day, and half were fixed factors that were held constant in the different versions of models. The random slopes models failed to converge and

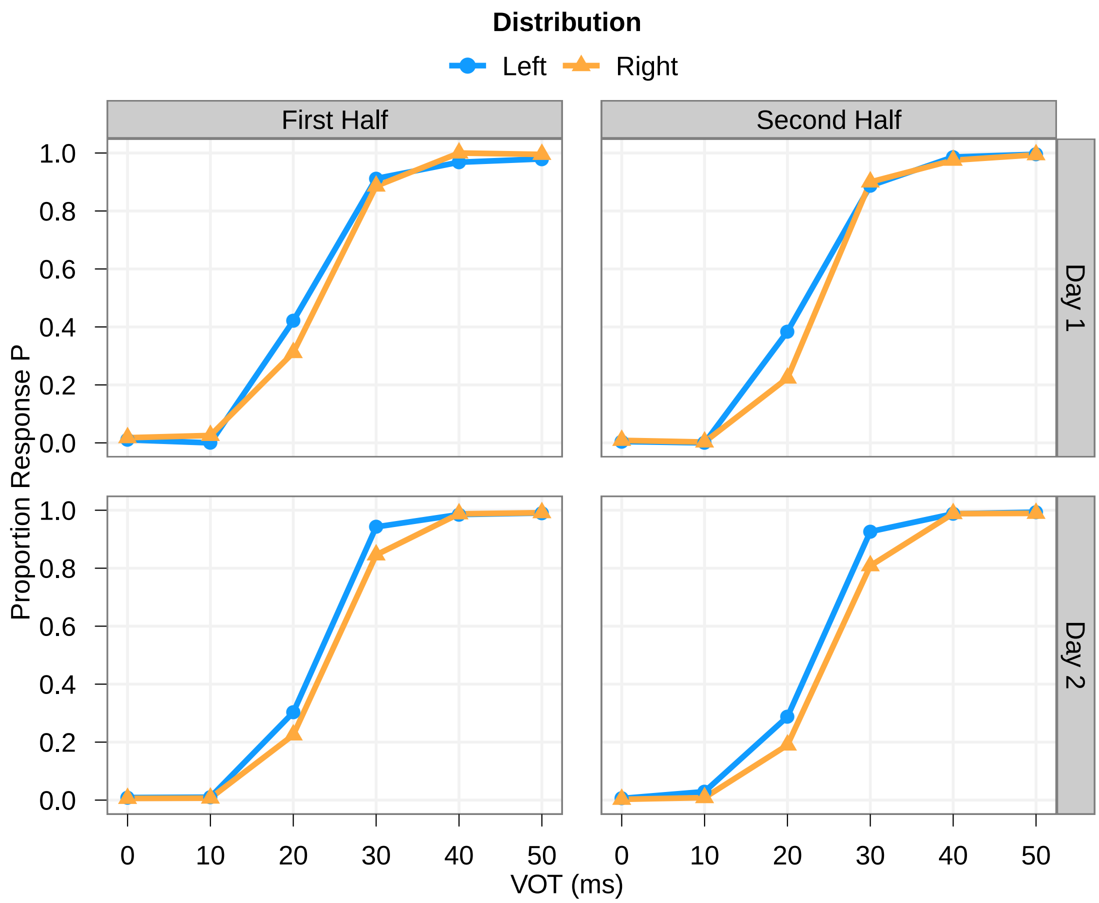

Figure 3.12: Experiment 4 distribution effect by experiment quarter.

including random intercepts for continuum improved the fit of the model compared to a version with random intercepts only for participants (χ^2(2)=246.39, p<.0001).

The full perceptual learning model is reported in Table 3.30. We are primarily interested in the main effect of distribution and the interactions involving distribution, day, and half, since these interactions show how the effect of distribution changed over the course of the experiment. All of the main effects were significant. Most critically, distribution (β=-1.48, p<.0001) affected responses as predicted: categorization boundaries for continua heard in the left-shifted distribution were to the left relative to those

Table 3.30: Experiment 4 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.35	0.30	1.2	>0.2
Distribution	−1.48	0.09	−16.8	**<.0001**
Day	−0.17	0.08	−2.1	**<.05**
Half	−0.24	0.08	−2.9	**<.01**
VOT Step	0.23	0.00	54.0	**<.0001**
Distribution x Day	−0.78	0.17	−4.7	**<.0001**
Distribution x Half	−0.45	0.17	−2.7	**<.01**
Day x Half	0.16	0.17	1.0	>0.3
Distribution x Step	−0.01	0.01	−1.5	>0.1
Day x Step	0.00	0.01	0.1	>0.9
Half x Step	0.01	0.01	1.6	>0.1
Distribution x Day x Half	0.58	0.33	1.7	>0.1
Distribution x Day x Step	−0.05	0.02	−3.5	**<.001**
Distribution x Half x Step	−0.04	0.02	−2.7	**<.01**
Day x Half x Step	−0.02	0.02	−1.4	>0.2
Distribution x Day x Half x Step	0.04	0.03	1.3	>0.2

Note: The maximum correlation among the fixed factors was r=.12, between distribution and step.

heard in the right-shifted distribution. Figure 3.12 shows the distribution effect in each experiment quarter. Distribution interacted with day (β=-.78, p<.0001) and half (β=-.45, p<.008), and marginally with both (β=.58, p<.09). These interactions suggest that participants learned the distributions over time since the boundary shifts were greater on the second day of the experiment and during the second half of each session. Overall, this analysis supports the idea that listeners can learn lexically-specific boundaries, since changes in lexical responding were consistent with the VOT distribution manipulations and increased exposure to these distributions. Since there were significant two- and three-way interactions between distribution, day, and half, we next ran simple effects analyses to determine when the effect of distribution was significant.

The simple effects models for both days are reported in Table 3.31. Just like in our analyses of phoneme specificity, on the first day there was a significant effect of distribution (β=-1.17, p<.0001), so participants were able to learn the distributions quickly. There was also an effect of half (β=-.35, p<.004) and an interaction between distribution and half (β=-.75, p<.002) showing that the distribution effect was larger in the second half of the first experiment session. On the second day there was a significant effect of distribution (β=-1.77, p<.0001) but the distribution by half interaction was not significant (β=-.14, p>.5). Further analyses of the first day (reported in Table 3.32) revealed effects of distribution for both halves of the session (First Half: β=-.81, p<.0001; Second Half: β=-1.44, p<.0001). These results indicate that listeners learned the distributions very quickly, as boundary shifts were present on both days and even during the first quarter of the experiment.

3.3.2.3 Vowel Effects

Since there were three pairs of continua that varied by vowel in this experiment, we also ran a secondary analysis to determine whether the distribution effect varied by vowel. Since the previous analysis showed similar distribution effects over the course

Table 3.31: Experiment 4 simple effects for days one and two.

		Coef β	SE(β)	**z**	**p**
Day 1	Intercept	0.46	0.35	1.3	>0.2
	Distribution	−1.17	0.13	−9.2	**<.0001**
	Half	−0.35	0.12	−2.9	**<.01**
	VOT Step	0.24	0.01	37.1	**<.0001**
	Distribution x Half	−0.75	0.24	−3.1	**<.01**
	Distribution x Step	0.01	0.01	1.3	>0.2
	Half x Step	0.02	0.01	2.0	**<.05**
	Distribution x Half x Step	−0.07	0.02	−2.9	**<.01**
Day 2	Intercept	0.25	0.26	1.0	>0.3
	Distribution	−1.77	0.12	−14.4	**<.0001**
	Half	−0.14	0.12	−1.2	>0.2
	VOT Step	0.22	0.01	39.5	**<.0001**
	Distribution x Half	−0.14	0.23	−0.6	>0.5
	Distribution x Step	−0.04	0.01	−3.4	**<.001**
	Half x Step	0.00	0.01	0.2	>0.9
	Distribution x Half x Step	−0.02	0.02	−0.9	>0.4

Note: The maximum correlation between fixed factors was r=.21 for the first day and r=.056 for the second day, both between distribution and step.

Table 3.32: Experiment 4 simple effects for each half of day one.

		Coef β	SE(β)	**z**	**p**
First Half	Intercept	0.67	0.42	1.6	>0.1
	Distribution	−0.81	0.17	−4.7	**<.0001**
	VOT Step	0.24	0.01	26.2	**<.0001**
	Distribution x Step	0.05	0.02	3.3	**<.001**
Second Half	Intercept	0.28	0.29	0.9	>0.3
	Distribution	−1.44	0.18	−7.8	**<.0001**
	VOT Step	0.24	0.01	26.4	**<.0001**
	Distribution x Step	−0.01	0.02	−0.8	>0.4

Note: The correlation between fixed factors was r=.31 for the first half and r=.18 for the second half.

of the experiment, we collapsed across the day and half factors in order to simplify the new model. In this model, reported Table 3.33, vowel was defined as a fixed factor with three levels (/i/, /ɪ/, and /ʌ/). Our primary interest was in vowel by distribution interactions since these would indicate a difference in the boundary shift effect for the different pairs of continua. There was a two-way interaction between distribution and vowel /ɪ/ (β=-.79, p<.0003), so we ran simple effects analyses to determine whether the effect of distribution was significant in each of the three vowel contexts.

The vowel simple effects models are reported together in Table 3.34 and responses by distribution for each vowel context are plotted in Figure 3.13. The effect of distribution was significant in all three vowel contexts (/i/: β=-1.02, p<.0001; /ɪ/: β=-1.90, p>.0001; /ʌ/: β=-1.09, p>.0001), indicating boundary differences in the predicted direction for continua with all three vowels. We were surprised that there was an effect of distribution for the words with an /i/ vowel since there is no visible difference between

Table 3.33: Experiment 4 vowel model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.16	0.40	0.4	>0.7
Distribution	−1.16	0.14	−8.1	**<.0001**
Vowel /ɪ/	−0.15	0.53	−0.3	>0.8
Vowel /ʌ/	0.67	0.53	1.3	>0.2
VOT Step	0.20	0.01	34.9	**<.0001**
Distribution x Vowel /ɪ/	−0.79	0.22	−3.7	**<.001**
Distribution x Vowel /ʌ/	0.01	0.21	0.0	>0.9
Distribution x Step	−0.02	0.01	−2.0	**<.05**
Vowel /ɪ/ x Step	0.05	0.01	4.8	**<.0001**
Vowel /ʌ/ x Step	0.03	0.01	2.8	**<.01**
Distribution x Vowel /ɪ/ x Step	0.00	0.02	−0.2	>0.9
Distribution x Vowel /ʌ/ x Step	0.05	0.02	2.7	**<.01**

Note: The maximum correlation between fixed factors was r=.22, between distribution and step.

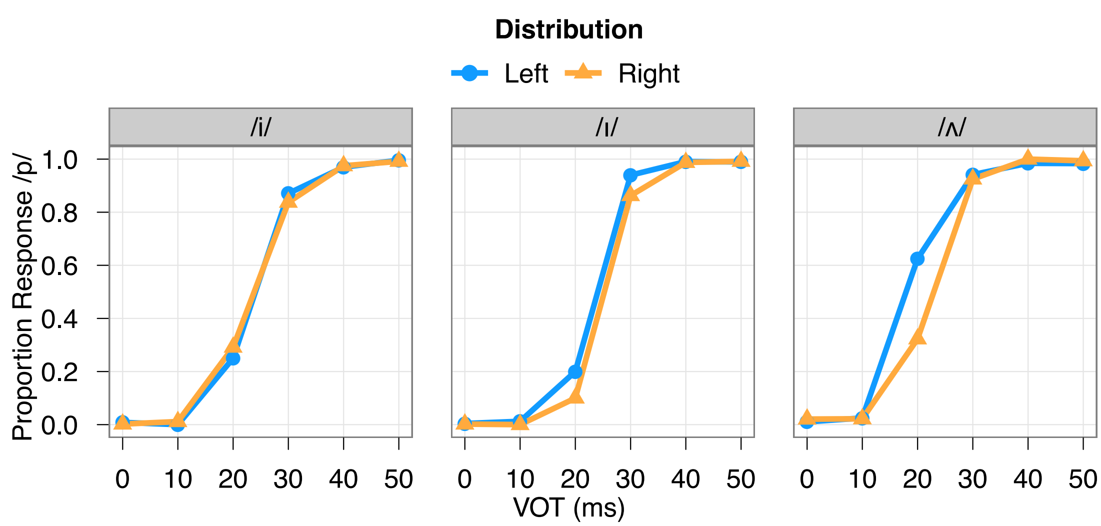

Figure 3.13: Experiment 4 distribution effect by stimulus vowel.

Table 3.34: Experiment 4 simple effects for vowels.

		Coef β	SE(β)	**z**	**p**
/i/	Intercept	0.17	0.20	0.8	>0.4
	Distribution	−1.02	0.13	−7.7	**<.0001**
	Step	0.21	0.01	33.5	**<.0001**
	Distribution x Step	−0.02	0.01	−2.0	**<.05**
/ɪ/	Intercept	0.00	0.16	0.0	>0.9
	Distribution	−1.90	0.16	−11.8	**<.0001**
	Step	0.25	0.01	28.8	**<.0001**
	Distribution x Step	−0.02	0.02	−1.4	>0.2
/ʌ/	Intercept	0.80	0.22	3.7	**<.001**
	Distribution	−1.09	0.15	−7.2	**<.0001**
	Step	0.22	0.01	31.7	**<.0001**
	Distribution x Step	0.03	0.01	2.1	**<.05**

Note: The correlation between fixed factors was r=.23 for the /i/ model, r=.14 for the /ɪ/ model, and r=.26 for the /u/ model.

the two distributions in Figure 3.13. However, when we plotted each of the two /i/ continua separately (shown in Figure 3.14), the difference between the two distributions was quite apparent for the *beak/peak* continuum, which explains the distribution effect in this vowel context.

3.3.3 Discussion

Experiment 4 tested whether word-specific perceptual learning is possible. We did this by manipulating the VOT distributions of word pairs with the same CV onsets, and examining how these distributions affected categorization. Participants showed

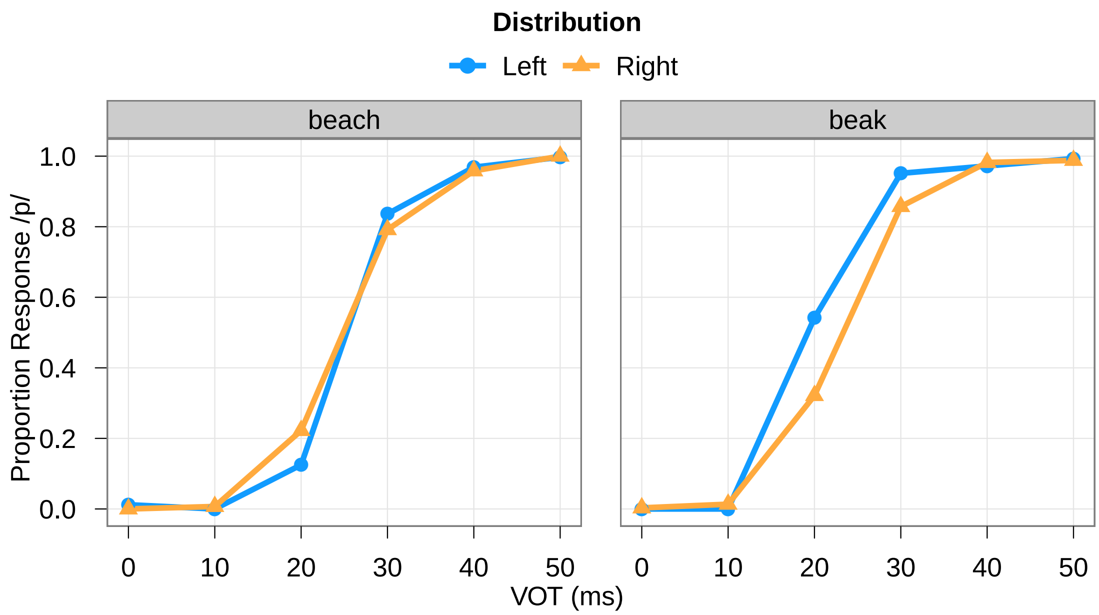

Figure 3.14: Experiment 4 distribution effects for /i/ continua.

categorization patterns consistent with the distributions they were exposed to for different words: left-shift words had more /p/ responses (a left-shifted boundary), and right-shift words had more /b/ responses (a right-shifted boundary). This pattern was stronger during the second half of each experiment session and during the second day overall, suggesting perceptual learning of the distributions over time. Since pairs of continua with matching CV onsets were randomly assigned to opposing distribution conditions for each participant, it is unlikely that the distribution effects were caused by prior variation in lexically-specific boundaries.

Our interpretation of the current results is not that listeners do not generalize across words, but that lexically-specific learning is also possible. Generalization is important for learning, and there have been many studies showing that listeners generalize perceptually learned boundary shifts to untrained words (Allen & Miller, 2004; McQueen et al., 2006; Maye, Aslin, & Tanenhaus, 2008; A. Hervais-Adelman et al., 2008; Sjerps &

McQueen, 2010). However, we find that under the right conditions, listeners can also learn lexically-specific boundaries.

3.4 General Discussion

The experiments in this chapter tested lexical and phoneme specificity in perceptual learning as well as generalization across words and onset places of articulation. The results showed evidence in favor of both generalization and specificity.

In Experiment 2 we found that listeners were able to learn multiple boundaries for the same feature contrast in different phonemic contexts–the boundaries appeared to be phoneme or place-specific, not generalized on the basis of the shared voicing feature. However, this specificity was limited to listeners who heard the bilabial onset continua shifted towards the left. Listeners who heard the coronal continua in the left distribution showed no boundary difference between the left- and right-shifted continua. This may have been because of secondary voicing cues present in our stimuli which made it difficult for participants to shift the coronal boundary to the left and the bilabial boundary to the right.

In Experiment 3 we found evidence supporting both generalization and specificity. Initially listeners generalized perceptually learned boundaries to new continua with both the trained and untrained onset place of articulation. After exposure to a different boundary for the second place of articulation, however, listeners used the new boundary for all of the continua with that place onset, and the initially trained boundary for the continua with the initially trained place. This suggests that listeners can learn place- or phoneme-specific boundaries, but generalize across different phonemic contexts when they lack more specific boundary information.

Finally, in Experiment 4 we found that listeners can learn lexically-specific boundaries as well. Learning these boundaries may even have been easier for listeners than learning phoneme-specific boundaries. Although this seems counterintuitive, the secondary voicing cues for the continua in the lexical-specificity experiment (which all shared the same onset place of articulation) should have been more similar than the secondary cues in the phoneme-specificity experiment (which had different onsets).

These results of these three experiments are somewhat different from other research on perceptual learning. Previous research has suggested that listeners generalize boundaries across different phonemes (Kraljic & Samuel, 2006; Theodore & Miller, 2010; Clarke & Luce, 2005) and across words (Allen & Miller, 2004; McQueen et al., 2006; Maye, Aslin, & Tanenhaus, 2008; Sjerps & McQueen, 2010). These results do not completely conflict with ours since we found evidence for generalization as well, and the ability of listeners to learn phoneme-specific and lexically-specific boundaries has not been previously examined. Thus, the existing literature on perceptual learning, in combination with our results, suggests that listeners can learn highly specific boundaries but also generalize learned boundaries across words or phonemic contexts.

As discussed in the introduction, the combination of specificity and generalization is not a pattern of results that many models of speech perception are prepared to handle. In general, theories that involve direct connections from acoustic information to words without intermediate levels (such as exemplar theory) are well-prepared to handle lexical-specificity but not generalization. Theories that involve abstract sub-lexical representations (such as TRACE and Merge) can handle generalization but not specificity.

The models that might allow listeners to use lexically-specific information but also generalize across words are quite variable. A dual-route model could allow for both generalization and specificity with one pathway directly from acoustics to words and another pathway through some level of abstraction (like features or phonemes). ART could also handle both generalization and specificity because it does not have defined levels or connections between levels, and can flexibly weight different types of information in response to task demands. Finally, C-CuRE could handle both generalization and specificity because it allows listeners to take lexically-specific information into account when it is available. While the initial bottom-up processing might be somewhat abstract, C-Cure could allow phoneme-specific or lexically-specific information to play a role in the relativization of cues, providing listeners with both the ability to make abstractions and take highly specific information into account.

Flexibility seems to be the key feature shared among models that might allow for both generalization and specificity. All of these models allow for flexible speech processing by way of multiple paths, re-weighting of information, or optional processing. The results of our experiment on lexically-specific perceptual learning suggest that listeners are sensitive to and take advantage of lexically-specific information when processing speech. Models of speech perception need to reflect listeners' use of this type of information, which may require increased flexibility.

CHAPTER 4
PERCEPTUAL LEARNING FOR MULTIPLE TALKERS

The experiments in this chapter address Aims 4 and 5: to test whether listeners spontaneously exhibit talker-specific perceptual learning in a task that does not emphasize talker identification, and to assess whether sequential versus simultaneous exposure to multiple talkers affects the degree of talker-specificity in learning.

According to exemplar theories of speech perception (e.g. Goldinger, 1996, 1998; Johnson, 1997), listeners store detailed representations of the input to which they are exposed, including indexical information. During speech recognition, both indexical and phonetic information is mapped onto existing exemplars, with no intermediation from sub-lexical units. This suggests that talker-specific perceptual learning should be easy. Parsing theories would also be able to account for talker-specific boundary learning by conditionalizing boundaries by talker. This would require storage of talker-specific information (e.g. the type of VOT values typically produced by a given talker). Prototype theories would find it more difficult to account for talker-specific learning, since the point of having prototypes is to eliminate the need for storage of more specific information.

It's not entirely clear how many of these models could account for both talker-specific learning and generalization across talkers, but exemplar models may be able to do so. While exemplar models have been criticized for positing storage of too many detailed exemplars, these models do not suppose unlimited storage. It is possible that listeners only store some number of the exemplars they experienced most recently. If this were the case, we would expect to see talker-specific learning when listeners are given mixed training on multiple talkers, but generalization across talkers occurs when they receive blocked training–exemplars from the more recently trained talker might overwrite those from previous talkers. Parsing theories might also predict differences in generalization and specificity based on training paradigm–listeners might do talker-specific learning when trained on two talkers simultaneously, but not bother storing talker-specific information when it is unclear that it is needed (leading to generalization across talkers). Similarly, the dual-route model could allow for both generalization and specificity by way of different processing routes—the phonological route could store general boundaries, while the exemplar route could handle talker-specific boundaries. Thus, Experiments 5 and 6 test whether listeners learn talker-specific boundaries when trained on two talkers at once (Experiment 5), but generalize across talkers when training is blocked by talker (Experiment 6).

4.1 Experiment 5

The aim of this experiment was to test whether listeners spontaneously exhibit talker-specific perceptual learning in a task that did not emphasize talker identification. Listeners were exposed to two talkers, one male and one female, with different VOT distributions. As in previous experiments their task was simply to click on the picture of the word that they heard on each trial, so they were not required to pay attention to talker differences in order to complete the task.

4.1.1 Method

4.1.1.1 Design

Experiment 5 used the same stimulus items as Experiments 1, listed again in Table 4.1. The design was also similar to Experiment 1 in that participants were exposed to shifted VOT distributions of male and female talkers. The main difference from the first experiment is that in Experiment 5, all participants were exposed to both talkers and both distributions instead of a single talker with one of the distributions. Trials from the two talkers were intermixed in a random order. The assignment of talker to distribution was counterbalanced across participants so that half of the participants heard the male

Table 4.1: Experiment 5 stimulus items.

/b/	/p/	/l/	/r/
beach	peach	lace	race
bees	peas	lake	rake
beak	peak	lei	ray
bit	pit	lock	rock
bin	pin	lamp	ramp
bill	pill	lane	rain

Note: These items are identical to those used in Experiment 1.

Table 4.2: Experiment 5 VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	4	36	70	36	4	4	36	70	36	4	0	0
Right-Shifted Distribution	0	0	4	36	70	36	4	4	36	70	36	4

talker with the left distribution while the other half of the participants heard the male talker with the right distribution. The female talker was always heard with the other distribution. The two distributions were centered at the same steps used as prototypes in Experiment 1, but because listeners heard both talkers they heard only half as many exemplars from each distribution. The numbers of exemplars presented at each VOT step are shown in Table 4.2. These distributions occurred across continua rather than within each continuum, which would have required many more trials.

Listeners heard a total of 300 experimental trials per day, 150 per talker. These were split evenly among the six continua so there were 25 experimental trials per continuum for each talker. There were also 25 filler trials per item-set for each talker, so there was an equal number of experimental and filler trials. Since the 25 filler trials for each talker and item-set could not be split evenly between /l/ and /r/ onset fillers, listeners who heard an extra /l/ filler for an item-set in the male talker's voice heard an extra /r/ filler in the same set with the female talker's voice. This kept the number of /l/ and /r/ onset filler trials even for each word and item-set (across talkers). Listeners heard a total of 600 trials per session and completed two sessions (approximately one week apart) for a total of 1200 trials.

4.1.1.2 Participants

Participants were 20 individuals from the University of Iowa community who participated in the study in exchange for course credit or a nominal payment. All were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. Three participants did not return for the second day of the study and were therefore excluded from analysis, leaving 17 participants who completed both experiment sessions.

4.1.1.3 Stimuli & Procedure

Stimuli were identical to those used in Experiment 1.

The procedure was identical to Experiment 1 as well. Participants were given no instructions concerning the two different talkers they heard during the experiment. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with the four images from an item-set (e.g. *beach, peach, lace,* and *race*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial. Although the Eye-Link II recorded participants' eye-movements throughout the experiment, we did not analyze the eye-movement data since the effects were large enough to observe in the mouse-click data.

4.1.2 Results

The structure of the results section mirrors that used for previous experiments, where each section addresses a different question. The first part of the results section addresses overall task performance. The second section assess perceptual learning of the two talker distributions over the course of the experiment. A boundary shift is predicted such that categorization data for continua from the talker trained on the left distribution will have a boundary shifted towards the left, while data for the other continua (from the talker trained on the right distribution) will show a boundary shifted towards the right. The final section of the results collapses across time to examine whether the distribution effect varied depending on shift condition (which talker was shifted to the left). Only steps -10 to 60 were analyzed in the mixed effects models since these were the steps shared between the two distributions.

4.1.2.1 Task Performance

Mouse-click responses were first examined to assess performance on the task. On critical trials where the stimulus began with a /b/ or /p/, participants clicked on a filler item on only .13% of the trials. We also found excellent performance on both endpoints: /b/s (-30 to -10 ms) averaged 99.2% correct and /p/s averaged 99.0% correct. Endpoint performance was also high for both talkers and distributions (see Table 4.3 for details). Three participants had endpoint performance below 75% correct and were thus

Table 4.3: Experiment 5 percept correct at endpoints for each talker.

	Male Talker		Female Talker	
Endpoint	Left	Right	Left	Right
/b/	99.5	100	99.8	97.5
/p/	97.5	99.6	100	99

excluded from further analysis.

4.1.2.2 Perceptual Learning of Talker-Specific Distributions

After establishing that endpoint performance was good, mouse-clicks across the continua were assessed in order to see whether participants learned boundary differences for the left- and right-shifted distributions. The response data are shown in Figure 4.1, with each quarter of the experiment in a separate panel. The boundary difference appears larger on the second half of both days.

To analyze these data we looked at the effect of distribution, day, and half (of each day). Response (/b/ or /p/, 0 and 1 respectively) was the dependent variable. VOT step, distribution (a within-participants variable), day, and half were fixed factors treated as in previous analyses. Participant and continuum were treated as random-effects. The random slopes models failed to converge and including random intercepts for continua as well as participants improved the fit of the model ($\chi^2(2)$=44.98, p<.0001), so this model was selected.

The model for distribution learning over time is reported in Table 4.4. There was a main effect of distribution (β=-.48, p<.02) indicating that the boundary for the talker with the left-shifted distribution was to the left of that for the talker with right-shifted distribution. There was also an interaction between distribution, day, and half (β=-2.50,

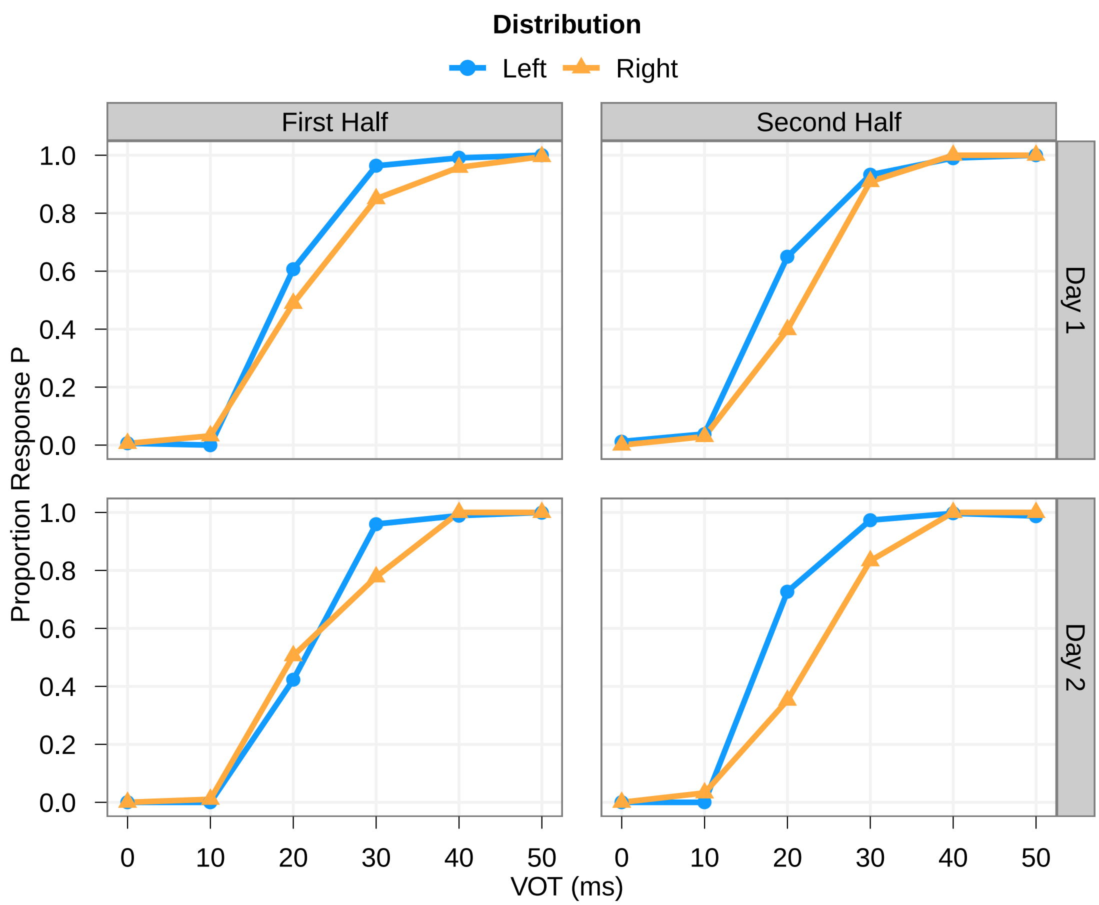

Figure 4.1: Experiment 5 distribution effect by experiment quarter.

Table 4.4: Experiment 5 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	1.35	0.28	4.8	**<.0001**
Distribution	−0.48	0.19	−2.5	**<.05**
Day	0.16	0.19	0.8	>0.4
Half	−0.31	0.19	−1.6	>0.1
VOT Step	0.30	0.01	25.5	**<.0001**
Distribution x Day	0.48	0.38	1.3	>0.2
Distribution x Half	−0.69	0.38	−1.8	>0.1
Day x Half	−0.61	0.38	−1.6	>0.1
Distribution x Step	0.04	0.02	1.9	>0.1
Day x Step	0.06	0.02	2.9	**<.01**
Half x Step	−0.03	0.02	−1.3	>0.2
Distribution x Day x Half	−2.50	0.76	−3.3	**<.01**
Distribution x Day x Step	0.02	0.04	0.3	>0.7
Distribution x Half x Step	−0.04	0.04	−0.8	>0.4
Day x Half x Step	−0.16	0.04	−3.6	**<.001**
Distribution x Day x Half x Step	−0.21	0.09	−2.3	**<.05**

Note: The maximum correlation among the fixed factors was r=.45, between distribution and step.

p<.001) and a marginal interaction between distribution and half (β=-.69, p<.08), which both suggested that the effect of distribution changed over the course of the experiment. This can be observed in Figure 4.1, where the identification functions do not appear equal in each quarter. Follow-up analyses (reported in Table 4.5) revealed an effect of distribution on the first day (β=-.75, p<.003), and no interaction with half (β=.51, p>.3), suggesting a fairly robust effect. On the second, there was no effect of distribution (β=-.29, p>.3). However, there was an effect of half (β=-.62, p<.04) and an interaction between distribution and half (β=-1.90, p<.002). Further analyses (reported in Table 4.6) suggested that this interaction was due to the fact that there was no effect of distribution in the first half of the second day (β=.66, p>.2), but a significant effect of distribution during the second half of the second day (β=-1.22, p<.0008). These results suggest that over the week between the two experiment sessions, participants did not retain the talker-specific boundaries learned on the first day. However, they were clearly able to re-learn the boundary shifts on the second day.

4.1.2.3 Talker-condition Effects

Since the assignment of talker to distribution was counterbalanced across participants, half of the participants heard the male talker with the left-shifted distribution and the other half heard the male talker with the right-shifted distribution. The next analysis collapses across the two days of the experiment to examine whether talker-condition affected boundary learning. The responses for each talker-condition are shown in Figure 4.2, where there appears to be a large boundary difference for the group with the female talker shifted to the left, and no difference for the group with the male talker shifted to the left. In the model, talker-condition was dummy-coded as 0 for the listeners who heard the male talker shifted to the left and 1 for the female talker shifted to the left. All listeners heard both talkers and distributions, but talker-condition was a between participants variable. The random slopes model failed to converge and including ran-

Table 4.5: Experiment 5 simple effects for days one and two.

		Coef β	SE(β)	**z**	**p**
Day 1	Intercept	1.30	0.34	3.9	**<.001**
	Distribution	−0.75	0.25	−3.0	**<.01**
	Half	0.01	0.25	0.1	>1
	VOT Step	0.27	0.01	20.3	**<.0001**
	Distribution x Half	0.51	0.50	1.0	>0.3
	Distribution x Step	0.03	0.03	1.3	>0.2
	Half x Step	0.05	0.03	2.1	**<.05**
	Distribution x Half x Step	0.06	0.05	1.2	>0.2
Day 2	Intercept	1.39	0.25	5.6	**<.0001**
	Distribution	−0.29	0.29	−1.0	>0.3
	Half	−0.62	0.29	−2.1	**<.05**
	VOT Step	0.32	0.02	17.0	**<.0001**
	Distribution x Half	−1.90	0.58	−3.3	**<.01**
	Distribution x Step	0.05	0.04	1.2	>0.2
	Half x Step	−0.11	0.04	−2.9	**<.01**
	Distribution x Half x Step	−0.14	0.07	−1.9	>0.1

Note: The maximum correlations among fixed factors were r=.39 for the first day and r=.45 for the second day, and were between distribution and step on both days.

Table 4.6: Experiment 5 simple effects for each half of day two.

		Coef β	SE(β)	z	p
First Half	Intercept	1.77	0.35	5.0	**<.0001**
	Distribution	0.66	0.46	1.4	>0.2
	VOT Step	0.39	0.03	11.4	**<.0001**
	Distribution x Step	0.12	0.07	1.7	>0.1
Second Half	Intercept	1.06	0.22	4.7	**<.0001**
	Distribution	−1.22	0.36	−3.4	**<.001**
	VOT Step	0.26	0.02	14.3	**<.0001**
	Distribution x Step	−0.03	0.04	−0.8	>0.4

Note: The correlations between fixed factors were r=.61 for the first half and r=.21 for the second half.

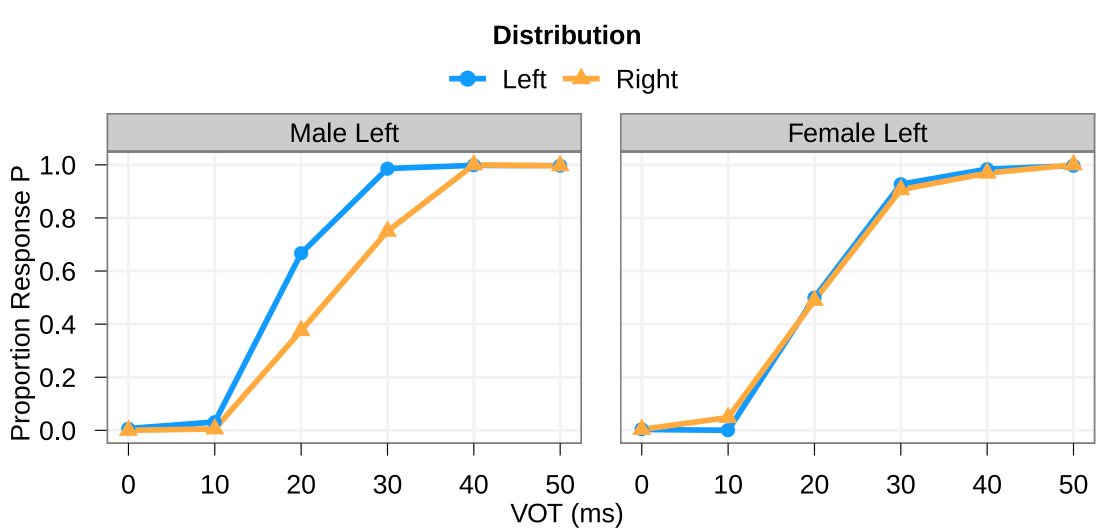

Figure 4.2: Experiment 5 distribution effect by talker-condition.

Table 4.7: Experiment 5 talker-condition model.

	Coef β	SE(β)	**z**	**p**
Intercept	1.36	0.28	4.9	**<.0001**
Distribution	−0.80	0.20	−4.0	**<.0001**
Talker-condition	0.42	0.37	1.1	>0.3
VOT Step	0.29	0.01	27.2	**<.0001**
Distribution x Condition	−1.44	0.39	−3.6	**<.001**
Distribution x Step	0.02	0.02	0.9	>0.3
Condition x Step	0.06	0.02	2.7	**<.01**
Distribution x Condition x Step	−0.01	0.04	−0.2	>0.8

Note: The maximum correlation among the fixed factors was r=.27, between distribution and step.

dom intercepts for continua as well as participants improved the model fit ($\chi^2(2)$=43.15, p<.0001), so this model was selected.

The talker-condition model, reported in Table 4.7, showed the same main effect of distribution (β=-.80, p<.0001) that was seen in the perceptual learning model. There was also an interaction between distribution and talker-condition (β=-1.44, p<.0003) showing that that the effect of distribution was greater for listeners who heard the female talker with the left distribution. Follow-up analyses on each talker (reported in Table 4.8) showed that there was an effect of distribution for the listeners who heard the female talker shifted left (β=-1.62, p<.0001), but no distribution effect for the group that heard the male talker shifted left (β=-.09, p>.7), suggesting that listeners only learned a boundary difference between the two talkers if the female talker was heard in the left distribution.

Table 4.8: Experiment 5 simple effects for talker-condition.

		Coef β	SE(β)	**z**	**p**
Female Talker Left	Intercept	1.54	0.37	4.2	**<.0001**
	Distribution	−1.62	0.32	−5.0	**<.0001**
	VOT Step	0.31	0.02	18.1	**<.0001**
	Distribution x Step	0.01	0.03	0.2	>0.8
Male Talker Left	Intercept	1.20	0.41	2.9	**<.01**
	Distribution	−0.09	0.24	−0.4	>0.7
	VOT Step	0.27	0.01	21.0	**<.0001**
	Distribution x Step	0.02	0.02	1.0	>0.3

Note: The correlation between the fixed factors was r=.15 for the female-left model and r=.48 for the male-left model.

4.1.3 Discussion

Experiment 5 aimed to test whether talker-specific perceptual learning is possible in a task that does not require talker identification, and when VOT was not one of the more prominent cues to talker identity. We did this by manipulating the VOT distributions of continua for the same words produced by two different talkers. Overall, participants showed boundary differences between the two distributions, but this effect seems to have been driven by the group that heard the female talker with the left distribution since there was no effect of distribution for the other half of the participants.

This result is similar to that found in Experiment 2, which tested place-specific perceptual learning. Listeners only showed boundary differences in the group that heard the /b/ words shifted to the left. The same explanation given also applies to this experiment: in retrospect, it seems likely that the auditory stimuli contained cues to voicing other than VOT (such as F1 and pitch) which made it difficult to shift the male continua to the left and female continua to the right. By training these shifted boundaries, for half of the participants we are effectively asking them to relearn secondary cues to voicing as opposite what they have previously known. Men have lower formant frequencies than women, and listeners associate lower F1 values with more voiced sounds (Summerfield, 1977). As a result, to shift the male talker's boundary to the left (so that there are more voiceless sounds along the continuum), listeners must learn that higher (instead of lower) F1 values go with voiced sounds. This may explain why listeners had trouble shifting the male talker's boundary to the left of the female's even though male talkers tend to have shorter VOTs (Swartz, 1992).

Although one group of participants did not learn separate boundaries for the two talkers, the other group did. This suggests that learning different voicing boundaries for different talkers (or perhaps different groups of talkers) is possible. Since this experiment trained listeners on boundary differences between a male and female talker, it is possible that listeners are learning about gender differences in voicing boundaries rather than talker-specific differences. In any case, listeners do not appear to be restricted to a single boundary. If listeners can learn multiple voicing boundaries, why then have previous studies found generalization across talkers? One possibility is that this is due to training listeners on each talker in a separate block instead of with intermixed trials. This hypothesis is addressed in Experiment 6.

4.2 Experiment 6

The aim of this experiment was to see if a blocked design would lead to generalization across talkers. After training on a single talker, listeners may generalize perceptually learned boundary shifts to a novel talker. Training on this second talker may then shift the boundary for both talkers, although if blocked training is not responsible for generalization across talkers, then it is possible that on the second day, listeners will show talker-specific boundaries after they have been trained on both talkers.

4.2.1 Methods

4.2.1.1 Design

Experiment 6 used the stimulus items from Experiments 1 and 5. Listeners heard distributions that varied by talker and were exposed to the distribution of only one talker per day (the training distribution). The distributions used in Experiment 6 are shown in Table 4.10. Like Experiments 1 and 5, the left- and right-shifted distributions have different ranges in this experiment. After exposure to the training talker for the day, listeners completed generalization trials (in a flat distribution) for the second talker. Trials for the two talkers are referred to as training and testing trials, but training remained implicit as it was in Experiment 1, and participants received no instructions that distinguished the generalization testing from the training trials.

Participants received 264 experimental training trials per day, divided equally between six continua. They received the same number of filler trials in the training talker's voice, also divided evenly between item-sets, and between /l/ and /r/ onset fillers for each item-set. The item-sets are listed in Table 4.9. In total participants completed 528 training trials during each of two experiment sessions.

Testing for generalization across talkers was limited to VOT steps -10 to 60 as in Experiment 3 (the phoneme generalization Experiment). Each of the 8 testing steps was repeated once (a uniform distribution) for each of the 6 continua, totaling 48 experimental testing trials per day. Listeners also heard an equal number of filler testing trials, again divided equally between item-sets and onset phoneme. Listeners completed a total of 96 testing trials per experiment session.

Talker voice and distribution shift direction on the first day were counterbalanced across participants, and the untrained talker and distribution were trained on the second day. The training talker from the first day became the tested talker on the second day.

Table 4.9: Experiment 6 stimulus items.

/b/	/p/	/l/	/r/
beach	peach	lace	race
bees	peas	lake	rake
beak	peak	lei	ray
bit	pit	lock	rock
bin	pin	lamp	ramp
bill	pill	lane	rain

Note: These items are identical to those in Experiments 1 and 5.

Table 4.10: Experiment 6 VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	4	31	62	31	4	4	31	62	31	4	0	0
Right-Shifted Distribution	0	0	4	31	62	31	4	4	31	62	31	4

4.2.1.2 Participants

Participants were 22 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay. All were monolingual native English talkers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. Six participants who did not complete both days were excluded from analysis, leaving 16 participants (4 in each condition) who completed both sessions of the experiment.

4.2.1.3 Stimuli and Procedure

Stimuli were re-used from Experiment 1 and were therefore identical to those described in Chapter 2.

The experiment procedure was also identical to that used in Experiment 1. Although this experiment added a testing component, testing trials were indistinguishable from training trials except by talker voice, and no new directions were given to participants before they began the testing trials. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with the four images from an item-set (e.g. *beach, peach, lace,* and *race*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial. Although participants' eye-movements were recording by the Eye-Link II, the effects were large enough to observe with the mouse-click data, so eye-movement data was not analyzed.

4.2.2 Results

The structure of the results section mirrors that used for previous experiments, where each section addresses a different question. The same mixed-effects modeling strategy is also employed. The first part of the results section addresses overall task performance. The second section assesses perceptual learning of the training distributions. A boundary shift is predicted such that categorization data for continua trained on the left distribution (in either the male or female talker's voice) will have a boundary shifted towards the left, while data for the other continua (those in the other talker's voice) will show a boundary shifted towards the right. The final two sections address generalization trials. We wanted to determine whether listeners would use the trained boundary for the generalization talker on each day. On the first day this was a new talker that hadn't been heard before, but on the second day this was the talker who the listener had heard a week ago with a different VOT distribution. Since these are very different trial-types we examined each day separately and asked whether boundaries varied between training and testing trials.

4.2.2.1 Task Performance

Mouse-click responses were first examined to assess overall performance on the task. Participants selected filler images on experimental trials (e.g. *lace* or *race* when the stimulus was *beach* or *peach*) for only .19% of the experimental training trials, which shows that they were paying attention to the task and not selecting images randomly. Performance on the unambiguous endpoints of the continua was also high. For the male talker, participants were 99.9% correct for both the /b/ (-30 to -10 ms) and /p/ (60 to 80 ms) endpoints. For the female talker, participants were 99.9% correct for the /b/ endpoint and 97.9% correct for the /p/ endpoint. The slightly lower accuracy at the female /p/ endpoint was driven by one participant who was at 70% correct for this endpoint. All other participants were above 98% percent correct at all the endpoints.

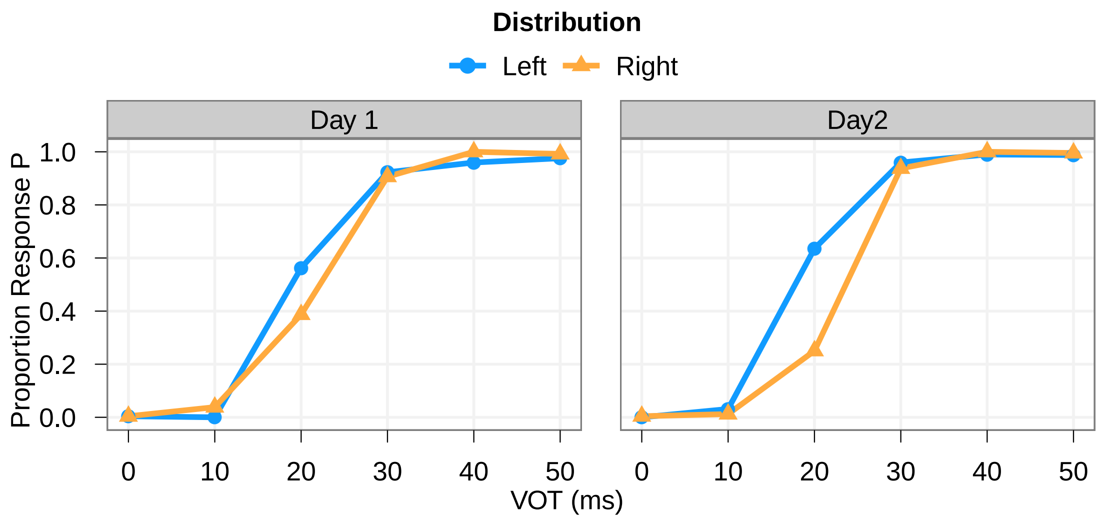

Figure 4.3: Experiment 6 training distribution by day. Participants in the left distribution on day 1 are in the right distribution on day 2.

4.2.2.2 Perceptual Learning for Training Distributions

Before assessing generalization to the different types of testing trials, we needed to establish that training was successful. To do this we first examined the effect of distribution on the two days of the experiment. Response (/b/ or /p/, coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution, and day were the fixed factors. Distribution was a within participants variable across the two training days, but between participants for the training on each day (since listeners were trained on only one talker and distribution each day). Participant and continuum were treated as random-effects. The random slopes models failed to converge and including random intercepts for continua as well as participants increased the model fit ($\chi^2(2)$=15.28, p<.0001), so this is the model we selected.

The model for Experiment 6 distribution learning is reported in Table 4.11. There was a main effect of distribution in the predicted direction (β=-1.01, p<.0001), indicating

Table 4.11: Experiment 6 perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.67	0.22	3.1	**<.01**
Distribution	−1.01	0.16	−6.3	**<.0001**
Day	0.04	0.16	0.3	>0.8
VOT Step	0.23	0.01	29.4	**<.0001**
Distribution x Day	−1.06	0.68	−1.6	>0.1
Distribution x Step	0.00	0.02	0.2	>0.8
Day x Step	0.04	0.02	2.6	**<.01**
Distribution x Day x Step	−0.04	0.03	−1.1	>0.3

Note: The maximum correlation among the fixed factors was r=.31, between distribution and step.

that participants had voicing boundaries shifted to the left for the talker heard with the left-distribution (relative to the talker heard with the right distribution). This is shown in Figure 4.3.

Since participants in Experiment 5 could successfully shift the female talker boundary to the left and the male talker boundary to the right, but not the opposite, we needed to examine the effect of talker-condition in this experiment as well. To do this we collapsed across day and half (the previous analysis showed not difference in the distribution effect over time) and included talker-condition as a new factor in the model. This between-participants variable was coded as 0 for participants who heard the male talker with the left distribution and 1 for participants who heard the female talker with the left distribution. The random slopes models failed to converge and including random intercepts for continua as well as participants improved the fit of the model ($\chi^2(2)$=14.77, p<.0001), so this is the model version we report.

Table 4.12: Experiment 6 talker-condition model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.69	0.23	3.0	**<.01**
Distribution	−0.95	0.16	−5.9	**<.0001**
Talker-condition	−0.04	0.37	−0.1	>0.9
VOT Step	0.24	0.01	28.9	**<.0001**
Distribution x Condition	0.65	0.32	2.0	**<.05**
Distribution x Step	0.01	0.02	0.8	>0.4
Condition x Step	0.02	0.02	1.2	>0.2
Distribution x Condition x Step	0.11	0.03	3.4	**<.001**

Note: The maximum correlation among the fixed factors was r=.36, between distribution and step.

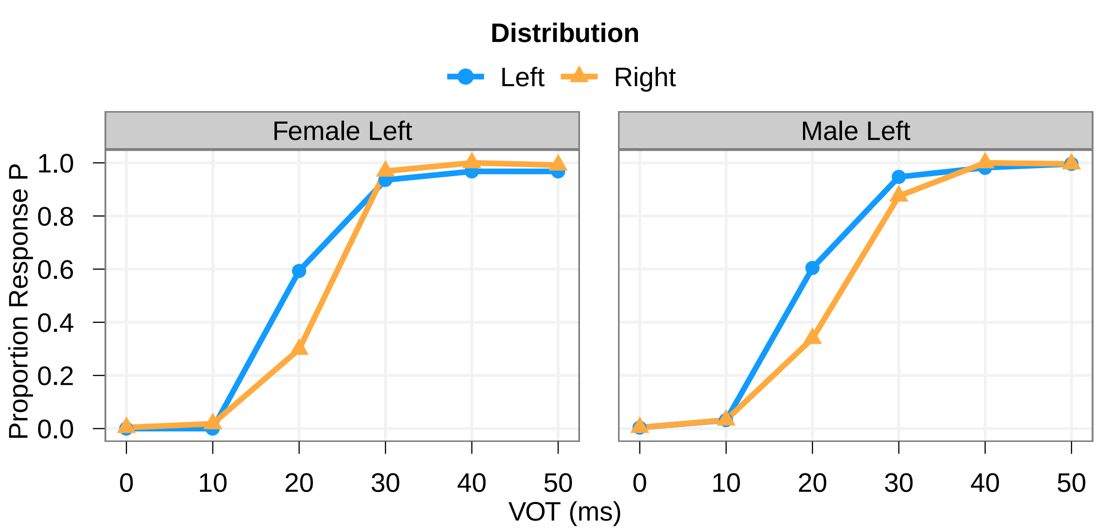

Figure 4.4: Experiment 6 training distribution by talker-condition.

Table 4.13: Experiment 6 simple effects for talker-conditions.

		Coef β	SE(β)	**z**	**p**
Male Talker Left	Intercept	0.72	0.37	2.0	>0.1
	Distribution	−1.36	0.23	−5.9	**<.0001**
	VOT Step	0.24	0.01	21.4	**<.0001**
	Distribution x Step	−0.04	0.02	−2.1	**<.05**
Female Talker Left	Intercept	0.70	0.51	1.4	>0.2
	Distribution	−0.75	0.25	−3.1	**<.01**
	VOT Step	0.27	0.01	19.1	**<.0001**
	Distribution x Step	0.07	0.03	2.8	**<.01**

Note: The correlations between the fixed factors were were r=.16 for the male-left model and r=.42 for the female-left model.

The talker-condition model for Experiment 6 is reported in Table 4.12. The model showed a main effect of distribution (β=-.95, p<.0001) and an interaction between distribution and talker-condition (β=.65, p<.05), suggesting that the boundary shift was larger for participants who heard the male talker shifted to the left. A plot of the response data (Figure 4.4) suggests that unlike the previous experiment, there was a distribution effect for both groups of participants. Follow-up analyses (reported in Table 4.13) confirmed this was the case: the effect of distribution was significant in both models (Male Left: β=-1.36, p<.0001; Female Left: β=-.75, p<.003), so participants in both talker-condition groups were able to learn boundary shifts based on the distributions they were exposed to for each talker.

4.2.2.3 Generalization Trials on Day 1

The generalization trials from the first day were used to test whether listeners generalize trained voicing boundaries to novel talkers. We compared training and test-

Table 4.14: Experiment 6 day 1 generalization model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.72	0.29	2.5	**<.05**
Training Distribution	−0.57	0.45	−1.3	>0.2
Trial-type	0.17	0.21	0.8	>0.4
VOT Step	0.23	0.01	24.8	**<.0001**
Distribution x Trial-type	0.15	0.41	0.4	>0.7
Distribution x Step	0.00	0.02	0.0	>1
Trial-type x Step	0.03	0.02	1.4	>0.2
Distribution x Trial-type x Step	−0.06	0.05	−1.2	>0.2

Note: The maximum correlation among the fixed factors was r=.13, between training distribution and step.

ing trials from the first day using models similar to the previous analyses, although in this analysis distribution is a between-participants variable since we are only examining one day. Trial-type was coded as dummy variable with 0 for training trials and 1 for testing trials, and then centered. The random slopes models failed to converge and including random intercepts for continua as well as participants improved the fit of the model (χ^2(2)=21.19, p<.0001), so this is the model we report.

The day 1 generalization comparison model is shown in Table 4.14. There was no effect of distribution (β=-.57, p>.2), which we found surprising, although we thought it might be because training distribution for each day was between participants, and since we were including participants as a random factor, all the variance due to distribution would be attributed to participant. When we ran a version of the model without participants as a random effect (Table 4.15), we did see an effect of training distribution (β=-.39, p<.03). Critically, there was no effect of trial-type (train vs. test) in either

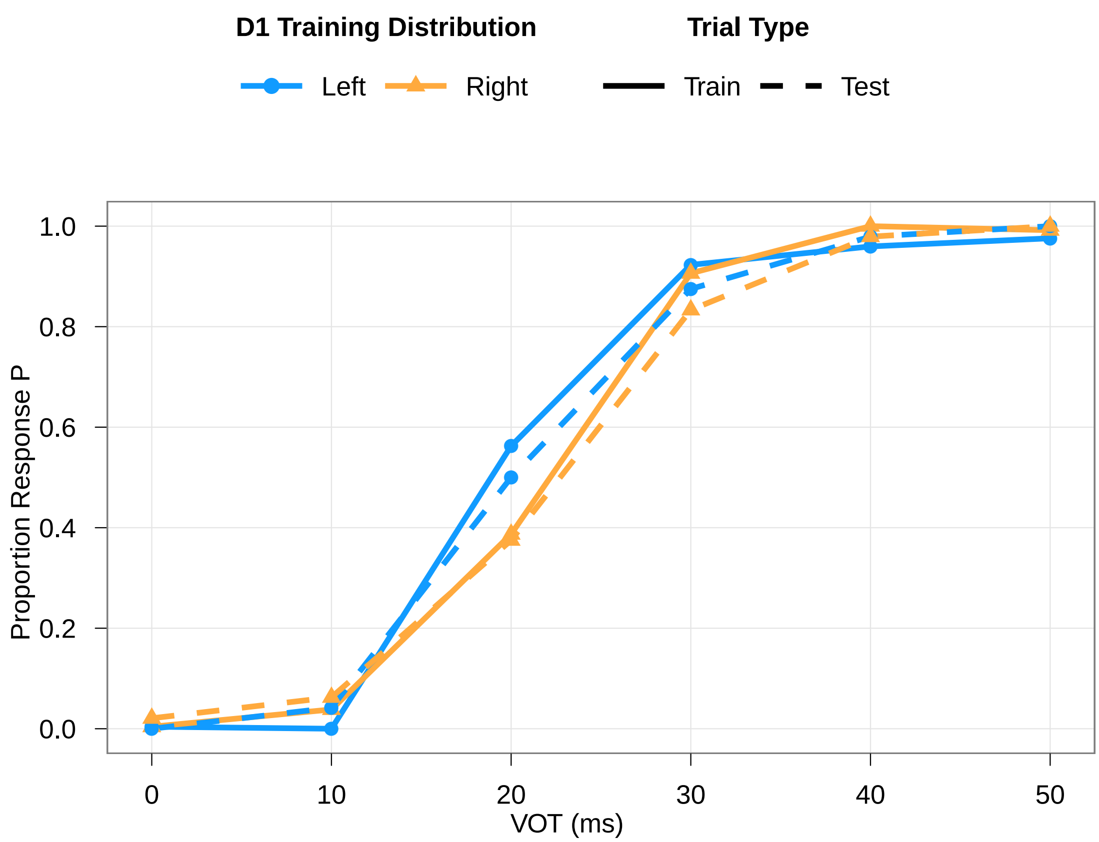

Figure 4.5: Experiment 6 day 1 training and testing trials.

Table 4.15: Experiment 6 day 1 generalization model without random participants.

	Coef β	SE(β)	**z**	**p**
Intercept	0.63	0.18	3.4	**<.001**
Training Distribution	−0.39	0.18	−2.3	**<.05**
Trial-type	0.17	0.20	0.9	>0.4
VOT Step	0.21	0.01	25.3	**<.0001**
Distribution x Trial-type	0.06	0.39	0.2	>0.9
Distribution x Step	0.02	0.02	0.9	>0.3
Trial-type x Step	0.03	0.02	1.2	>0.2
Distribution x Trial-type x Step	−0.05	0.04	−1.1	>0.3

Note: The maximum correlation between fixed factors was r=.4, between distribution and step.

model (Version1: β=.17, p>.8; Version 2: β=.17, p>.9), which indicates that participants did not respond differently to training and testing trials–they had the same boundary for the training and testing talkers on the first day. This suggests that listeners generalize learned voicing boundaries to novel talkers.

4.2.2.4 Generalization Trials on Day 2

The models of generalization on day 2 had the same structure as those for day 1. Because we coded talker-condition according to the training distribution on day 1, the distribution effect in the model (if any) is expected to have the opposite direction from that observed for the day 1 models. That is, we have typically expected to see a negative β for distribution because left is coded as 0 and right as 1. On the second day, since we are coding distribution according to day 1 condition, day 2 distributions are coded as 0 for right and 1 for left, so a positive β will indicate a distribution shift in the direction of the day 2 training distributions. The model with random slopes for participants and random

Table 4.16: Experiment 6 day 2 generalization model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.69	0.17	4.0	**<.0001**
Training Distribution	1.39	0.35	4.0	**<.0001**
Trial-type	−0.07	0.21	−0.4	>0.7
VOT Step	0.24	0.01	22.2	**<.0001**
Distribution x Trial-type	−0.97	0.41	−2.4	**<.05**
Distribution x Step	0.05	0.02	2.1	**<.05**
Trial-type x Step	−0.03	0.02	−1.2	>0.2
Distribution x Trial-type x Step	0.13	0.05	2.6	**<.01**

Note: The maximum correlation among the fixed factors was r=-.10, between training distribution and step.

intercepts for continua failed to converge, and including continuum as a random effect in addition to participants did not improve the fit of the model (χ^2(2)=48.21, p<.0001), so we report the model without the random effect of continuum.

The day 2 generalization model is shown in Table 4.16. There was a main effect of training distribution (β=1.39, p<.0001) but no effect of trial-type (β=-.07, p>.7). However, there was also an interaction between distribution and trial-type (β=-.97, p<.02) that we did not find on the first day. This can be observed in Figure 4.6, where the effect of trial-type appears to be reversed for the two training distribution conditions. The direction of the trial-type by distribution interaction is consistent with the training from day 1– that is, when listeners are re-tested on the day 1 training talker at the end of day 2, they appear to show boundaries consistent with their training on that same talker on day 1, rather than the other talker that was trained more recently. Follow-up analyses (Table 4.17) revealed that on day 2, the effect of trial-type was marginal for both groups (R:

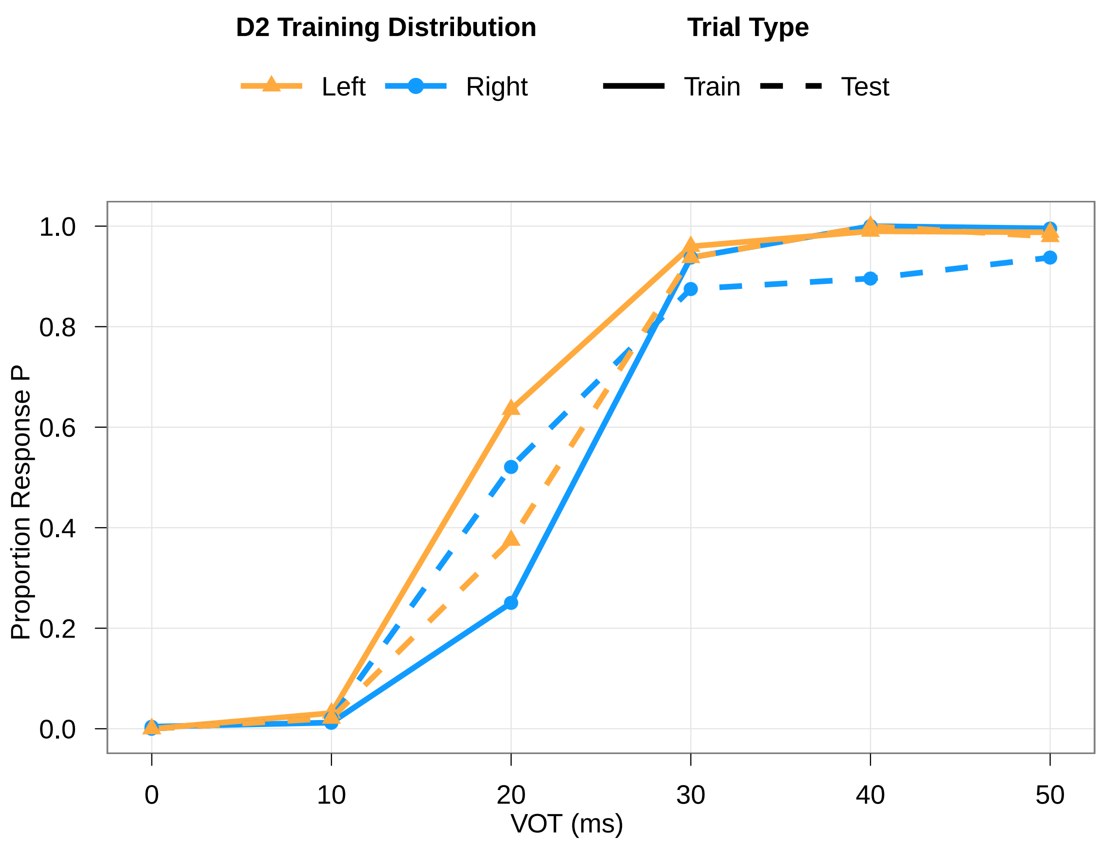

Figure 4.6: Experiment 6 day 2 training and testing trials.

Table 4.17: Experiment 6 simple effects for day 2 distribution groups.

		Coef β	SE(β)	**z**	**p**
Right-Shift on Day 2	Intercept	−0.07	0.22	−0.3	>0.8
	Trial-type	0.40	0.24	1.7	>0.1
	Step	0.24	0.02	13.9	**<.0001**
	Trial-type x Step	−0.09	0.02	−4.2	**<.0001**
Left-Shift on Day 2	Intercept	1.51	0.31	4.9	**<.0001**
	Trial-type	−0.57	0.34	−1.7	>0.1
	Step	0.26	0.02	14.3	**<.0001**
	Trial-type x Step	0.04	0.04	0.8	>0.4

Note: The correlations between the fixed factors were r=-.4 for the right-shifted model and r=-.03 for the left-shifted model.

β=.40, p<.09; L:β=-.57, p<.1). In both cases the trend was in the direction predicted by the day 1 training trials.

Therefore, our final analysis examined whether the generalization trials from day 2 had the same boundary as the training trials from day 1. The model with random slopes for participants and random intercepts for continua failed to converge, and including random intercepts for continua in addition to participants improved the fit of the model (χ^2(2)=8.96, p<.003), so this is the model we report.

The day 1 training and day 2 testing comparison model is shown in Table 4.18 and the data are plotted in Figure 4.7. There were no main effects of training distribution or trial-type, but there was an interaction between distribution and trial-type (β=1.29, p<.002). Follow-up analyses (Table 4.19) showed an effect of trial-type for the left-shifted listeners only (L: β=-.67, p<.006; R: Lβ=-.51, p>.1). This suggests that listeners in the group trained on the right-shifted distribution on the first day may have formed a more

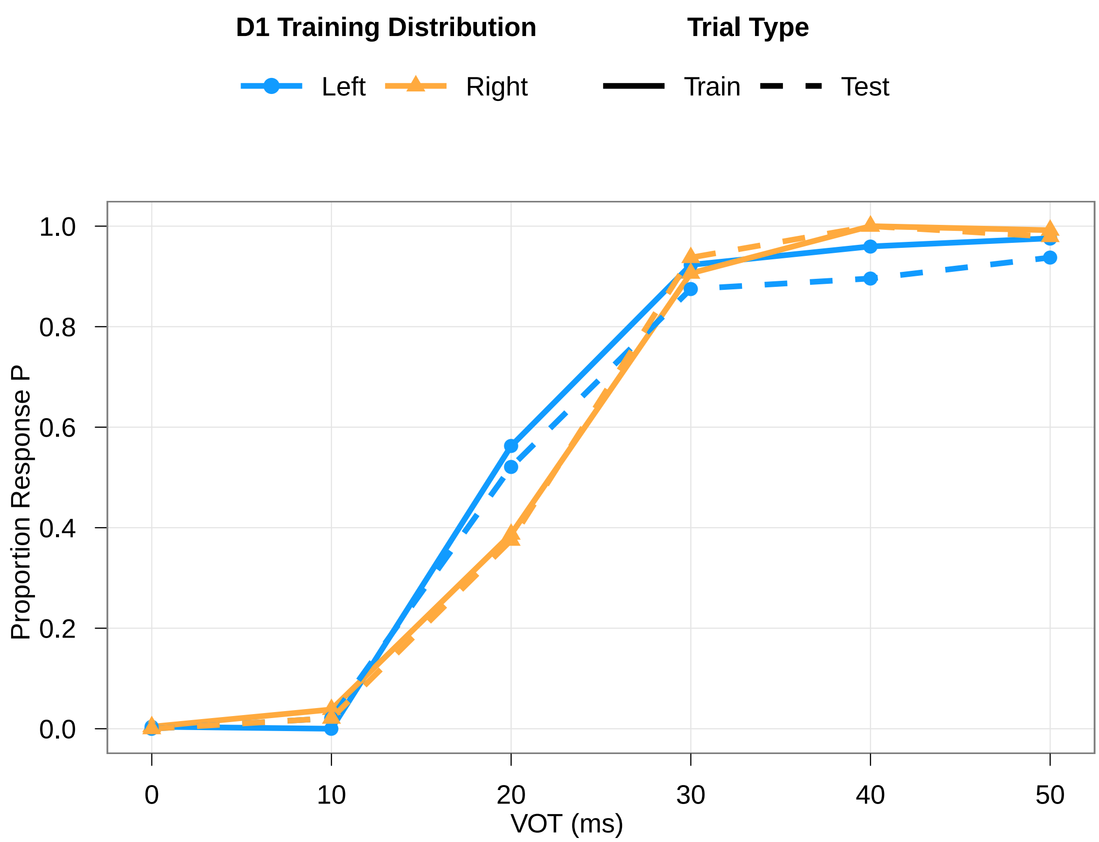

Figure 4.7: Experiment 6 day 1 training and day 2 testing trials.

Table 4.18: Experiment 6 day 1 train and day 2 generalization model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.73	0.30	2.5	**<.05**
Training Distribution	−0.44	0.53	−0.8	>0.4
Trial-type	−0.03	0.20	−0.2	>0.9
VOT Step	0.23	0.01	24.5	**<.0001**
Distribution x Trial-type	1.29	0.41	3.1	**<.01**
Distribution x Step	0.03	0.02	1.7	>0.1
Trial-type x Step	0.01	0.02	0.4	>0.7
Distribution x Trial-type x Step	0.13	0.05	2.8	**<.01**

Note: The maximum correlation among the fixed factors was r=.13, between training distribution and step.

robust representation of that distribution, since they showed the same boundary when re-tested on that speaker at the end of the second day.

4.2.3 Discussion

In Experiment 6 we aimed to test whether a blocked training design would lead to generalization across talkers. We found that after a day of training on a single talker and distribution, listeners showed no difference in categorization boundary between the trained talker and a second talker introduced to test for generalization at the end of the session. Similarly, on the second day listeners were able to learn new boundaries for the second talker, but did not use the same boundary when categorizing speech from the talker trained in the first day (with a different distribution). Instead, participants showed categorization boundaries more consistent with their training from the first day, so the intervening training on the new talker did not overwrite what they had learned about the first talker. This suggests that it is not blocked training that leads to

Table 4.19: Experiment 6 simple effects for day 1 training and day 2 testing comparison.

		Coef β	SE(β)	z	p
Right-Shift on Day 1	Intercept	0.43	0.30	1.5	>0.1
	Trial-type	0.51	0.32	1.6	>0.1
	Step	0.23	0.02	14.6	**<.0001**
	Trial-type x Step	0.07	0.04	1.7	>0.1
Left-Shift on Day 1	Intercept	1.06	0.44	2.4	**<.05**
	Trial-type	−0.67	0.24	−2.8	**<.01**
	Step	0.22	0.01	17.6	**<.0001**
	Trial-type x Step	−0.06	0.02	−2.9	**<.01**

Note: The correlation between the fixed factors was r=-.33 for the right-shifted model and r=.018 for the left-shifted model.

generalization across talkers, but rather a lack of talker-specific information. Perhaps previous experiments that have shown voicing boundary generalization across talkers have not provided listeners with enough training or the right kind of training necessary for forming talker-specific representations.

4.3 General Discussion

The experiments in this chapter tested talker-specificity and generalization in perceptual learning. Like the experiments in Chapter 3, we found evidence for both generalization and specificity.

In Experiment 5 we found that participants were able to learn talker-specific VOT distributions, but not when these distributions did not conflict with secondary voicing cues (e.g. F1). That is, listeners learned talker-specific boundaries when the female-talker had VOTs shifted towards the voiced end of the continuum, but not when the male talker was shifted in that direction. While we discuss this as talker-specific learning, it also possible that this learning is not truly talker-specific, but applies to groups of talkers (e.g. male and female). This still suggests that listeners can learn multiple boundaries for a contrast, and do so automatically in a task that does not require them to do so. Critically, this is difficult to accommodate in models in which talker compensation occurs prior to recognizing a phoneme or feature of some kind.

In Experiment 6 we found that listeners generalize learned voicing boundaries to novel talkers but retain talker-specific boundaries once they are learned. That is, listeners generalized boundaries learned on day 1 to a novel talker, but on day 2, learned a new boundary for the novel talker. When re-tested on the initial talker, boundaries were more consistent with the training from that talker rather than the more recently trained talker. This suggests that the more recent training on the second talker did not overwrite the initial training for the original talker, and that listeners can retain talker-specific boundaries even after prolonged exposure to other talkers.

The results of these experiments share some similarities with previous studies on perceptual learning, but also some differences. While talker-specific learning for voicing boundaries has been shown using tasks that require talker identification (Allen & Miller, 2004; Theodore & Miller, 2010), perceptual learning studies using the lexical-feedback paradigm have shown that voicing boundaries are generalized across talkers (Kraljic & Samuel, 2006, 2007). Perhaps listeners in these studies did not receive enough exposure to the initially trained talker to form a robust, talker-specific voicing boundary.

As discussed previously, the combination of specificity and generalization is a more complicated pattern than those predicted by most models of speech perception, and one that many models may not be prepared to handle. Theories that posit abstract representations of speech input (such as TRACE and Merge) are well-equipped to handle generalization across talkers, but not talker-specificity. Theories that posit veridical storage of speech input (such as exemplar theory) are well-equipped to handle talker-specificity, but the degree to which listeners should generalize across talkers is somewhat unclear. It is unlikely that these models could accommodate both generalization and talker-specificity for the same two talkers, which was the pattern of results that we observed.

Both parsing and dual-route models may be able to handle this pattern of results in different ways. Dual-route models could allow for both generalization and specificity with one pathway involving abstract talker-specific representations, and another pathway without this type of abstraction. It is logical that the abstraction pathway would involve abstraction across different types of variability (words, place of articulation, talkers), and the other direct pathway would be an exemplar-type route. This second pathway could accommodate talker-specific learning via direct connections between cues and words.

Parsing models could show the same pattern if listeners could conditionalize voicing categorization on talker when talker-specific representations were available. When such representations were not available (as with a novel talker), boundaries would generalize across talkers. Talker-specific representations could be stored in feedback connections between talkers and individual cues, and listeners could use these representations to re-code cue values on the basis of individual talkers. When encountering a novel talker such talker-specific representations would not be available and listeners would generalize on the basis of the category.

Either way, the results of our experiments on talker-specificity and generalization across talkers suggest that listeners are sensitive to and take advantage of differences between talkers when processing speech. Many models of speech perception do not reflect this sensitivity and should strive to incorporate the use of talker-specific information in the future.

CHAPTER 5
GENERAL DISCUSSION

The experiments in this dissertation had four specific aims. In the first part of this discussion I will review those aims and our results. In the next section I will address the shortcomings of our data. In the third section I will discuss the ways in which our results conflict with previous studies on perceptual learning, and possible reasons for our results being different. Next I will talk briefly about how our unsupervised learning paradigm may be related to developmental processes. Finally, and most importantly, I will discuss the implications of our results for theories and models of speech perception.

5.1 Specific Aims and Results

1) To test whether perceptual learning in speech is phoneme-specific. In Experiment 2, we found that listeners are able to learn multiple voicing boundaries for different pairs of phonemic contrasts relying on the same feature contrast. This supports phoneme-specificity in perceptual learning. For example, listeners can learn that the boundary for /b/ to /p/ continua was shifted to the left of that for /d/ to /t/ continua. The opposite shifts did not appear to be learnable, possibly because of secondary voicing cues that we did not (and could not) manipulate—the lower F2 of a /b/ and higher F2 of a /d/ are inherent to place of articulation. Listeners may not be able to learn all boundary shifts equally well because they conditionalize on secondary cues. Experiment 3 also provides support for phoneme-specific learning–boundaries for generalization trials on day 2 were consistent with the boundaries trained for each place of articulation.

2) To assess the degree of generalization to different phonemic contexts that rely on the same feature contrast. In Experiment 3, we found that listeners generalize voicing boundaries to untrained continua with the same onset as the trained continua. For example, a boundary learned in the context of *beak/peak* will generalize to *beach/peach*. Listeners can also generalize learned boundaries to continua with different onsets (like *dart/tart*), but only do so if they have not already learned a different boundary for other continua (like *dime/time*) with this onset. If listeners have experience with a boundary for a particular phonemic contrast, they retain that boundary and apply it to other continua with matching onsets. Generalization occurs when listeners do not have this type of experience.

3) To test whether perceptual learning can be specific to particular words. Experiment 4 found that listeners can learn different voicing boundaries for continua with the same CV onset. For example, the boundary for *beach/peach* can be shifted to the left while the boundary for *beak/peak* is shifted to the right. This suggests there is some degree of lexical-specificity in perceptual learning.

4) To test whether spontaneous talker-specific perceptual learning can be observed task that does not emphasize talker identification. Listeners in Experiments 5 and 6 learned different voicing boundaries for multiple talkers even though they were not given explicit instructions to pay attention to talker differences. In Experiment 5 exposure to the two talkers was simultaneous (with trials for both talkers randomly mixed on both days), while in Experiment 6 exposure to the two talkers was blocked by day, with listeners hearing a single talker and distribution on each day. In both cases, listeners were able to learn talker-specific boundaries (although in Experiment 6 they generalized these boundaries when they had experience with only a single talker), and the talker-specific learning occurred even though our unsupervised perceptual learning paradigm did not require listeners to use multiple boundaries.

5) To test whether sequential versus simultaneous talker training affects the degree of talker-specificity in learning. Generalization in Experiment 6 was related to previous experience with a talker, not to training structure. Listeners with no exposure to a new talker generalized a learned boundary to that talker, but listeners with previous experience with a talker retained that talker's boundary even after extensive exposure to a second talker. The blocked training in Experiment 6 appeared to be more powerful than the simultaneous training in Experiment 5, as it overcame the effect of secondary cues (F0 and F1).

5.2 Shortcomings in the Data

While the global pattern just described largely supports our hypotheses (and challenges current models of speech perception) there were some more idiosyncratic results. The best example of this was in Experiment 2 on phoneme-specific learning. Listeners were able to learn different boundary shifts for continua beginning with bilabial onsets and coronal onsets (both distinguished by voicing), but the participants who learned the left-shifted boundary for labial items (and the right boundary for coronal items) were responsible for effect–the other group showed no boundary difference between the two distributions. A likely explanation for this is that our continua contained secondary cues to voicing that indicated a boundary more on the left for the bilabial continua, and on the right for the coronal continua. These cues are inherent to place of articulation: bilabials, for example, have lower F2 onset transitions (a cue to voicing), while coronals have higher F2 onsets (a cue to voicelessness). Thus, in some conditions we were asking listeners to learn to use F2 in a completely reversed way (higher F2s/coronals leads to more voiced sounds). When asking participants to learn such conflicting distributions they were able to shift the boundaries only enough that they matched the boundaries instead of reversing them. This alone is intriguing as in typical, untrained, performance, listeners show a *d/t* boundary to the right of the *b/p* boundary (Sawusch & Pisoni, 1974). However, it also suggests that perceptual learning may not always overcome inherent differences in the stimuli on which the new boundaries may be conditionalized.

An alternative possibility is that listeners have an innate or a priori constraint that does not allow coronal voicing boundaries to be shifted to the left of bilabial voicing boundaries. We prefer the previous account to the constraint-based account because it provides a more explanatory mechanism. If there were to be such a constraint, where would it come from? If the constraint were learned, would this be particularly different from the previous account, which relies on learned cues to voicing? In fact, our secondary cue account could serve this same stabilizing role, without the need to resort to innate constraints on learning.

Similarly, in the Experiment 5 (the talker-specific learning experiment), we found that listeners only learned two different boundaries when the female talker was heard with the left distribution and the male with the right. The same explanations can also account for these results. Males have lower F1 and F0 values (cues to voicing) than females. This might make it difficult to shift male talker voicing boundaries such that more VOT values are identified as voiceless while shifting female voicing boundaries in the opposite direction. Intriguingly, in Experiment 6, in which listeners were trained on the two distributions and talkers on separate days, both groups of participants were able to learn the two different boundaries. This suggests that blocked training may allow listeners to more effectively track the distributions for individual talkers. A similar effect has been observed for listeners learning Mandarin tone contrasts, where at least some listeners benefit from reduced trial-by-trial variability (Perrachione, Lee, Ha, & Wong, 2011). Perhaps more broadly, Experiment 6 makes it clear that listeners can learn more difficult boundary shifts with sufficient or more effective training. This makes an innate constraint-based account of their inability to learn boundary shifts in particular directions less plausible. This could only account for our results of listeners can un-learn constraints, which would make it difficult to distinguish between the learned cues to voicing and constraint-based accounts of the data.

While interleaved training typically leads to more robust learning in other domains, perhaps it is worse in this case because blocking allows listeners to implicitly down-weight the irrelevant secondary cues which are more salient during interleaved training. Our suggestion that learning is better when exposure to the two talkers is blocked relies on the learnability of of the boundary shifts for both the male-left and female-left conditions. While blocking may facilitate acquisition of difficult boundary shifts, it may not have a positive effect on the magnitude of the boundary shifts observed–we have not done a direct comparison of the distribution effect in the two different experiments, but contextual interference predicts that the shifts should be smaller in the blocked condition.

At a broader level, the presence of these secondary cues (which may exert an important limitation on learning) is not just an artifact of stimulus construction–coronals have higher F2s than labials and men have lower F1s and F0s than women. This has important consequences for the types and directions of boundary shifts that are likely to be learnable, both in laboratory experiments and more typical real-world settings.

While we have referred to talker-specific learning throughout our discussion of our results, our data do not distinguish between boundaries that apply to specific subgroups of talkers or truly talker-specific boundaries. Since we had a male and female talker, it is impossible to know from our experiments if listeners can only learn one boundary that applies to all female talkers and another that applies to all male talkers, or it they can learn multiple boundaries for talkers of the same gender. Other experiments suggest that listeners can do talker-specific boundary learning for talkers of the same gender (Allen & Miller, 2004; Theodore & Miller, 2010), but this remains to be demonstrated with our paradigm. We predict that listeners can learn talker-specific boundaries in this paradigm, although they may require additional training. Blocking the exposure to the two talkers may be more effective, as we saw in Experiment 6. Crucially, however, this does not undermine the broader theoretical implications of these experiments: although listeners may have learned boundary shifts for groups of talkers rather than individual talkers, this still supports our assertion that listeners employ specific boundaries (either talker-specific or group-specific) in certain circumstances, and more general boundaries in others.

5.3 Conflicts with Previous Studies

A number of our results conflict with those of previous experiments on perceptual learning in speech perception. These studies have almost entirely used the lexically driven learning paradigm in which an ambiguous sound is disambiguated by lexical context, thereby shifting the category boundary so that the previously ambiguous sound is included in the appropriate category. In these studies, lexical feedback provides an implicit error signal for what category the ambiguous sound must belong to. While we have referred to this as supervised learning, it is not a typical form of supervised learning since listeners receive no external error signal telling them whether their decisions were correct or incorrect. However, lexically supervised learning still contrasts with our unsupervised or distributional paradigm, in which tracking how frequently different VOT values occur is sufficient to drive a boundary shift. The conflicts between our results and previous studies suggest that previously hypothesized limits on perceptual learning in speech perception may need to be relaxed.

First, in Experiments 2 and 3 we found that listeners were able to learn different boundaries for two different types of stop-voicing continua (e.g. coronal and labial onsets), while previous studies have shown only generalization across these types of continua. For example, Kraljic and Samuel (2006) found that listeners exposed to ambiguous /d/ or /t/ sounds also generalized learned boundary shifts to a /b/ to /p/ continuum. This is consistent with our results in Experiment 3: on the first day of the experiment, listeners generalized the coronal or labial boundaries they learned to the untrained continua types. However, when listeners are trained on two different boundaries, we find that they can learn both. This may be a limitation of what previous studies have tested rather than a limitation of the lexical-feedback paradigm–there's no reason to think that listeners trained in the lexical-feedback paradigm used by Kraljic and Samuel (2006) might not also be able to learn more specific boundaries if they had exposure to multiple boundaries shifted in opposite directions.

Secondly, we found talker-specific perceptual learning for stop consonant voicing boundaries, even when exposure to the two talkers was blocked. Previous studies have shown generalization across talkers (Kraljic & Samuel, 2006, 2007). Since these studies also used talkers of opposite genders, the difference in results is unlikely to be due to higher similarity between talkers in the previous experiments. Kraljic and Samuel (2006) suggested that listeners learn talker-specific boundaries for spectral contrasts (like fricative voicing) and generalize across talkers for temporal contrasts (like stop consonant voicing, for which VOT is the primary cue). We failed to find any support for this pattern of results since our listeners were able to learn talker-specific boundaries for stop-voicing contrasts as well. Since unsupervised learning is typically considered to be less powerful than supervised learning, it is surprising that we observed talker-specific learning in our unsupervised paradigm, while Kraljic and Samuel (2007) did not find talker-specific learning using lexically-supervised learning. As listeners in the Kraljic and Samuel (2007) study were trained on different boundaries for the two talkers, it cannot be a complete lack of experience with a talker that leads to generalization across talkers. However, listeners in our study had much more experience with each talker, which may have led to more robust representations of each talker's boundary, despite the differences in training paradigm.

Finally, we found perceptual learning for shifted boundaries at the onset of words, while other studies have failed to show perceptual learning for boundary shifts when the critical segments are heard at word onset. Jesse and McQueen (2011) used the lexical-feedback paradigm and compared learning when the ambiguous segments were heard word-finally (as in Norris et al. (2003)) or word-initially. While listeners showed evidence of boundary shifts when the ambiguous segments were heard word-finally, they did not have shifted boundaries when ambiguous segments were heard word-initially. In contrast, we found that unsupervised perceptual learning is capable of driving boundary shifts for word-initial segments. This is one way that our unsupervised learning paradigm may be more powerful than the lexical-feedback paradigm.

It is likely that our paradigm is more powerful due to the increased number of exposure trials that listeners get. The lexical-feedback paradigm provides more information in a single trial than our distributional learning paradigm, which requires listeners to track distributions over a number of trials, but even so the difference in trial numbers is striking. Participants in Kraljic and Samuel (2007) heard 20 critical trials for each talker, but participants in Experiment 6 (the blocked talker experiment) heard 264 critical trials per talker. Our increased exposure time may also be responsible for driving boundary shifts for segments at word-onset, although the lack of effect in the lexical-feedback paradigm could also be due to the source of information driving the boundary shift. Lexical items are more active by the time listeners reach the final phoneme than at word onset. When the ambiguous segment is encountered word initially, listeners may struggle to retain the information until they have built up the lexical activation necessary to disambiguate the segment. When the ambiguous sound is word final, this is not a problem. This provides an alternative explanation for why listeners in our distributional learning paradigm are able to learn boundary shifts for word-initial sounds: these boundary shifts are not driven by lexical feedback. While it remains to be seen whether some types of learning we found can also be supported by other training paradigms, another advantage to unsupervised learning is that it maps onto what children are likely to do during development.

5.4 Learning and Development

In our studies, participants learned phoneme boundaries by tracking the distributional statistics of the input. This fits well with theories of learning in infancy (e.g. Maye et al., 2002; Maye, Weiss, & Aslin, 2008; McMurray, Aslin, & Toscano, 2009; Kuhl, 2004). Very young infants lack the lexical or phonotactic knowledge that is needed as an error signal for supervised learning, and yet it is clear that significant speech development occurs over the first year of life (e.g. Werker & Tees, 1984; Kuhl et al., 2006; Werker & Curtin, 2005). This makes unsupervised learning an attractive candidate mechanism for infant speech category development: infants could track how frequently different values along a given cue dimension occur (e.g. VOT values), and then use the clusters to form categories (e.g. voiced and voiceless). If infants do use this mechanism to acquire sound categories then the same mechanism might also be used to adjust these categories during adulthood.

Although we do not have evidence that the perceptual learning observed in our experiments is driven by the same learning mechanism used by infants, we do not have a reason to believe that it must be a different mechanism either. The simplest explanation is that listeners take advantage of all sources of information that are available to them. For infants, this may mean relying on distributional statistics. For adults, this mechanism may continue to be available in addition to others that become available as listeners acquire more knowledge necessary to take advantage of other types of information.

Although adults may use learning mechanisms beyond those available to infants, our studies suggest that unsupervised learning allows adults, and likely infants as well, to track highly specific distributional information. For example, infants may be able to track distributions specific to both talkers and lexical items, something that has not yet been tested during infancy. However, if infants are tracking such highly conditionalized statistics, it is not clear how they could be developing robust context-invariant phonological categories using this mechanism alone, as a number of researchers have suggested (Werker & Curtin, 2005; McMurray et al., 2009). On the other hand, perhaps infants do not have or need such categories, and rely on an exemplar-type route to speech processing. Perhaps the phonological route does not develop until later, possibly when children are learning to read.

5.5 Implications for Theories of Speech Perception

Most importantly, these experiments on both specificity and generalization of perceptual learning are critical to our understanding of the types of abstract units and the levels and pathways of processing in the speech perception system. Previous perceptual learning studies have generalization across factors like talker and phoneme and have used this to argue in favor of abstract units, suggesting that if listeners generalize a boundary shift along a *b/p* continuum to a *d/t* continuum as well, they must have adjusted the mappings between VOT and voicing features rather than specific phonemes (McQueen et al., 2006; Cutler, 2010; Cutler et al., 2010). Tests of specificity, however, provide a stronger test of abstraction, in that phoneme- or talker-specific learning should not be possible if speech perception is mediated by units of this sort. Thus, we should not interpret generalization as indicating that more specific learning is impossible.

Models of speech perception posit categorization at different points in processing, making the issue of abstraction contentious. Prototype models like TRACE and MERGE have abstract units that allow for categorization of speech information before it reaches the lexicon. As a result, if a listener learns that the *b/p* boundary should be shifted left in the context of *beach/peach*, they will have to generalize this to all other words (since this learning occurs in the mapping between VOT and phonemes). In contrast, exemplar models have no abstraction, even at the level of the lexicon. As a result, each word can be mapped quite independently to continuous speech cues like VOT, allowing for a high degree of specificity. By testing the level of specificity with which listeners are able to learn different voicing boundaries, we have sought to determine the level of abstraction found within the perceptual system.

The experiments presented here found evidence in favor of both specificity and generalization in perceptual learning. Listeners were able to learn voicing boundaries that are specific to 1) both talkers (or perhaps categories of talkers), 2) both places of articulation, and 3) even individual lexical items. However, at the same time, they also generalized boundaries to new talkers and new places of articulation (and by extension, new words). This generalization occurred when listeners did not have prior exposure to the stimuli that informed them of a different boundary location.

Since models of speech perception typically have abstract units or do not, this combination of specificity and generalization is not a pattern of results that many models are prepared to handle. Models that are able to accommodate these results do so by providing multiple ways of mapping low-level cues (like VOT) onto higher-level elements (such as words). This can be done with a variety of different mechanisms.

ART (Grossberg et al., 1997; Grossberg, 2003; Goldinger & Azuma, 2003) is one model that should be able to handle both talker-specific and lexically-specific information. One of its unique characteristics is that it has no defined units or connections. During speech processing, the perceptual system eventually reaches a state of resonance for a particular unit (which could be a phoneme, syllable, or word). Task demands can change the relative weighting of top-down and bottom-up information, thereby affecting the size of the units able to achieve resonance. In tasks demanding lexical-specificity, for example, lexically specific information should play a larger role and allow listeners to make categorization judgments specific to individual lexical contrasts. In other situations, features or phonemes might play a larger role, leading to generalization. Although

ART should be able to account for a wide variety of results in different circumstances, it may also be more flexible than human listeners. It is difficult to know whether it would show our pattern of results without running simulations.

Parsing models like C-CuRE (McMurray & Jongman, 2011; Cole et al., 2010; McMurray et al., 2011) can accommodate both specificity and generalization because they allow listeners to parse information with different sources of variability depending on experience. For example, listeners exposed to a single talker may adjust their categorization boundaries based on global characteristics like the task context, since this situation provides no evidence that talker-specific representations are necessary. If a second talker is introduced, then generalization will occur. Listeners exposed to multiple talkers, however, may track talker-specific characteristics. This might allow listeners to conditionalize cue values based on talker-specific representations, leading to talker-specific categorization judgments. Critically, these talker-specific representations would be stored in the feedback connections between talkers and individual cues (used to recode cue-values on the basis of talker), allowing the model to also use (and generalize) a single bottom-up category. While this architecture should be able to account for our pattern of results, this implementation requires us to make some specific architectural assumptions that were not a part of C-CuRE's original instantiation, which left C-CuRE's architecture unspecified.

A dual-route model is transparent in the way it provides multiple pathways for mapping cues to categories, which is an advantage it has over the other models with alternative mechanisms. Like TRACE, there is one pathway from cues to words that makes use of sub-lexical abstract units, a sort of "phonological" pathway. However, unlike TRACE the dual-route model has an additional pathway directly from cues to words with no intermediate abstraction. The first pathway allows for generalization on the basis of abstract units. Feature-like units, for example, would allow listeners to generalize a shifted *b/p/* boundary to other contrasts that rely on the same cues, like the *d/t* or *g/k* voicing contrasts. This pathway would also allow listeners to generalize across talkers, assuming the abstract units were not talker-specific. The second pathway directly from cues to words would allow for more specific boundary learning, like talker-specific or lexically specific boundaries. Having two routes may be advantageous because a direct route might be faster or more efficient for typical speech perception, but a phonological route would allow for processing of non-words, metalinguistic task performance, and generalization across talkers when encountering a novel talker.

While a number of approaches to speech perception can potentially accommodate our findings that listeners can show both specificity and generalization for talkers and lexical contrasts, other models will require modification. The different mechanisms used by the dual-route model, parsing models, and ART may have varying degrees of compatibility with other models of speech perception. For example, TRACE and MERGE could be fairly compatible with the dual-route model since they already have defined pathways and units that could instantiate one of the two routes—these defined units and connections are incompatible with ART, which is designed to function without specific units and connections.

The high degree of learning specificity encountered in these experiments suggests a certain arbitrariness to the learning, which is to say that this learning may not be specific to speech perception. Perhaps listeners could learn different distributions of associated with different colored backgrounds instead of different talkers. Although colored backgrounds are not related to speech stimuli in any obvious way, the speech perception system would still need to have some representation of voicing for these colored backgrounds to have an effect on voicing continua. While the results of these experiments may be due to a more general learning process not specific to speech, they clearly indicate what representations must be present in the speech processing stream.

5.6 Conclusion

While there have been many previous studies on the generalization of perceptual learning, the inverse issue of specificity has been relatively neglected. The experiments presented here contribute to filling this gap. While generalization is an important aspect of perceptual learning for speech, a degree of specificity is also critical—without specificity, listeners would be unable to cope with individual variability or adapt to multiple dialects or accents. We have found evidence that listeners can learn voicing boundary shifts that apply only to specific talkers (or groups of talkers), places of articulation, and lexical contrasts, as well as generalizing across both talkers and place. This suggests that the speech processing system has mechanisms that allow for both specificity and generalization of learning, a feature shared by only a small subset of models. Speech is so noisy and context dependent that perhaps the only way to reliably perceive speech is for listeners to retain flexibility while simultaneously being able to generalize as the situation demands.

APPENDIX A
EXPERIMENT 2A

The experiment reported in this Appendix was the first attempt to get at the question better answered by Experiment 2. Experiment 2A was meant to test phoneme-specific boundary learning and was run before Experiment 2. As is reported below, the results were inconclusive. There was little indication of learning but we thought it possible that this was because the learning context emphasized only voicing and not place of articulation (the other feature that distinguished the two VOT distributions). The experiment was re-designed (Experiment 2) to provide listeners with a reason to pay attention to place as well as voicing.

Experiment 2A addresses Aim 1: to test phoneme-specific perceptual learning.The Introduction and Chapter 3 discuss this possibility and the theoretical implications of this type of learning in more detail. The goal was to expose listeners to conflicting distribution shifts in different phonemic contexts relying on the same feature contrast. For example, voicing is used to distinguish between /b/ and /p/ as well as /d/ and /t/. We wanted to see if listeners could shift the boundary between the bilabial sounds in one direction and the boundary between the coronals in the opposite direction. If listeners can learn different category boundaries for these two contrasts, it would indicate that these boundaries are not being learned at the feature level.

A.1 Method

A.1.1 Design

Experiment 2A tested whether listeners can show evidence of phoneme-specific boundary shifts driven by unsupervised perceptual learning. Voicing continua with one onset place (e.g. bilabial) were shifted to the left and voicing continua with another onset place (e.g. coronal) were shifted to the right. The shift direction for each place

Table A.1: Experiment 2A stimulus items.

/b/ or /d/	/p/ or /t/	/l/	/r/
beach	peach	lace	race
bin	pin	lake	rake
bug	pug	lei	ray
bear	pear	lock	rock
dart	tart	lamp	ramp
deer	tear	lane	rain
dime	time	leaf	reef
dent	tent	list	wrist

of articulation was counterbalanced across participants. We used four continua (each with a different vowel) for each place of articulation. The words at the endpoints of these continua are shown in Table A.1 along with the filler items that were used in this experiment. Each set of experimental and filler items was kept constant across participants and only the male talker was used for this experiment.

For the distribution shift manipulation, the left and right distributions were centered at the same steps used in Experiments 1 and 2, but the number of repetitions at each step was modified to accommodate the different number of continua used. Like the distributions in Experiment 2, both of the distributions extended the full width of the VOT range. The number of exemplars presented at each VOT step for the two distributions is shown in Table A.2. Distributions were applied across the four continua at each place of articulation.

Participants completed a total of 320 critical trials per day, 160 critical trials in each distributions/place of articulation. These were split evenly between the four con-

Table A.2: Experiment 2A VOT distributions.

VOT Step	-30	-20	-10	0	10	20	30	40	50	60	70	80
Left-Shifted Distribution	3	18	34	18	4	4	18	34	18	3	3	3
Right-Shifted Distribution	3	3	3	18	34	18	4	4	18	34	18	3

tinua at each place, so there were 40 critical trials per continuum in each distribution. There were an equal number of filler trials for each item set that were split evenly between the /l/ and /r/ fillers. The 320 experimental and 320 filler trails totaled 640 trials per session. The experiment has two sessions held approximately a week apart (1280 trials all together). Listeners heard the continua shifted in the same direction on the second day as they had on the first.

A.1.2 Participants

Participants were 23 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay. All participants were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. 21 participants completed both sessions of the study, and 2 participants completed only one session. These 2 participants were excluded from analysis.

A.1.3 Stimuli

A.1.3.1 Auditory Stimuli

Auditory stimuli consisted of eight twelve-step VOT continua ranging from -30 to 80ms. Four continua were from /b/ to /p/ and the remaining four were from /d/ to /t/. Four of the continua were used in previously reported experiments. The *beach/peach* and *bin/pin* continua were used in Experiment 1, and the *dart/tart* and *deer/tear* con-

Table A.3: Experiment 2A bilabial VOT measurements.

beach/peach	bin/pin	bear/bear	bug/pug
-32	-29	-31	-32
-23	-19	-22	-24
-12	-10	-13	-8
0	0	0	0
10	9	7	9
21	20	18	19
31	30	29	29
41	40	40	39
51	50	50	50
61	60	60	60
71	71	70	70
80	80	80	80

tinua were used in Experiment 2. The remaining continua were new and were created in the same manner as the other continua, by cross-splicing recordings of natural speech. The recording and cross-splicing methods used to create the stimuli are described in Chapter 2. The same male speaker was recorded for the new continua. The recordings were made in the same location as the other continua and we tried to match the recording levels as closely as possible. VOT measurements for completed continua are listed in Tables A.3 and A.4.

The stimuli were piloted by six lab members using the same procedure as the stimuli in Experiment 1. On each pilot trial participants used a key press to identify the given stimulus as beginning with /b/ or /p/ for the bilabial continua, or /d/ or /t/ for

Table A.4: Experiment 2A coronal VOT measurements.

dart/tart	deer/tear	dent/tent	dime/time
-30	-30	-32	-33
-17	-20	-24	-23
-9	-12	-15	-9
0	0	0	0
10	10	10	10
20	20	20	20
30	30	30	30
40	40	40	40
50	50	50	50
60	60	60	60
70	70	70	70
80	80	80	80

the coronal continua. The two types of continua were piloted in separate blocks. Each of the stimulus items was repeated three times. At the voiced endpoint, listeners correctly identified the stimulus as a /b/ or /d/ on 100% of the trials. At the voiceless endpoint they identified the stimulus as a /p/ or /t/ on 98.3% of the trials. The point at which the identification functions crossed 50% was between steps 6 and 7 for all of the bilabial continua, and between 6 and 8 for the coronal continua.[1]

A.1.3.2 Visual Stimuli

Pictures representing each item listed in Table A.1 were constructed using the same picture norming technique described for Experiment 1. Images that were included in previous study were re-used here. The final images were approved by the author and thesis supervisor.

A.1.3.3 Procedure

The procedure was identical to that used in Experiment 1. An Eye-Link-II head-mounted eye-tracker was calibrated to each participant before the beginning of the experiment. Following calibration, participants read the instructions for the experiment and practiced the drift-correct procedure.

On each trial, participants were presented with four images from an item set, two experimental and two filler (e.g. *dent, tent, list,* and *wrist*). They saw one image in each corner and a red dot in the middle of the screen. After 500ms the dot turned blue. When participants clicked on the blue dot they heard the auditory stimulus for the trial over headphones. They clicked on the corresponding image and moved on to the next trial. Participants' eye-movements were recorded by the Eye-Link II but eye-movement data was not analyzed.

[1] The *dart/tart* and *dime/time* continua crossed the 50% point between steps 6 and 7 and the *deer/tear* and *dent/tent* continua crossed the 50% point between steps 7 and 8.

A.2 Results

A boundary shift is predicted such that categorization data for words trained on the left distribution (either bilabial or coronal onset) will have a boundary shifted towards the left while data for the other words (those trained on the right distribution) will show a boundary shifted towards the right.

The structure of the results section mirrors that used for previous experiments, where each section addresses a different question. The same mixed-effects modeling strategy is also employed. The first part of the results section addresses overall task performance and the second section assess perceptual learning of shifted boundaries over the course of the experiment.

A.2.1 Task Performance

Mouse-click responses were first examined to assess overall performance on the task. On experimental trials where the stimulus began with a /b/ or /p/, participants clicked on a filler item on only .13% of the trials, which indicates that they were paying attention to both the auditory and visual stimuli. Performance on the unambiguous endpoints of the continua was also high. On the voiced side (-30 to -10 ms), participants selected the /b/ image for 99.8% of the bilabial trials and the /d/ image for 99.7% of the coronal trials. On the voiceless side (60 to 80 ms), participants selected the /p/ image for 99.9% of the bilabial trials and the /t/ image for 97.9% of the coronal trials.

A.2.2 Perceptual Learning

The most critical analysis concerns learning of the shifted distributions. Response (voiced or voiceless, coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution, and day were fixed factors coded and centered as in previous experiments. Participant and continuum were treated as random-effects. The random slopes models failed to converge and adding random intercepts for continua as well im-

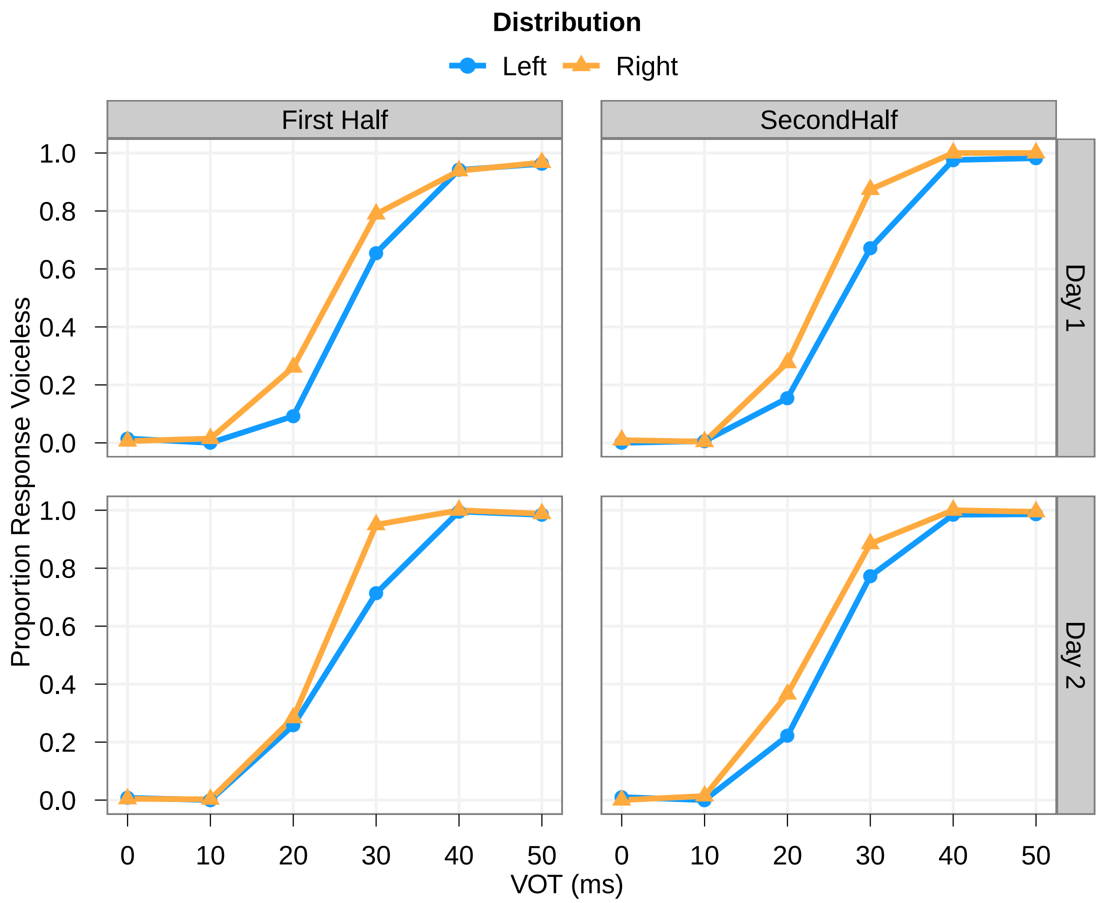

Figure A.1: Experiment 2A distribution effect.

proved the model fit (χ^2(2)=254.21, p<.0001), so we selected this model.

The perceptual learning model with the best fit to the data is reported in full in Table A.5. The model showed a main effect of distribution (β=-.40, p<.004) in the direction predicted by the distribution manipulation. There was also a distribution by half interaction (β=.56, p<.01) indicating that the distribution effect was larger in the second halves of the experiment sessions. While these effects were promising, we were concerned that they did not match our plots of the data, which seemed to show a distribution effect in the wrong direction (Figure A.1).

Table A.5: Experiment 2A perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.12	0.39	0.3	>0.8
Distribution	−0.40	0.14	−3.0	**<.01**
Day	0.33	0.11	3.1	**<.01**
Half	0.35	0.11	3.2	**<.01**
VOT Step	0.25	0.01	39.2	**<.0001**
Distribution x Day	−0.02	0.22	−0.1	>0.9
Distribution x Half	0.56	0.22	2.6	**<.01**
Day x Half	−0.30	0.22	−1.4	>0.2
Distribution x Step	0.04	0.01	2.9	**<.01**
Day x Step	0.03	0.01	3.0	**<.01**
Half x Step	0.04	0.01	3.6	**<.001**
Distribution x Day x Half	−0.01	0.43	0.0	>1
Distribution x Day x Step	0.02	0.02	0.7	>0.5
Distribution x Half x Step	0.06	0.02	2.4	**<.05**
Day x Half x Step	−0.09	0.02	−3.7	**<.001**
Distribution x Day x Half x Step	−0.01	0.05	−0.3	>0.8

Note: The maximum correlation between the fixed factors was r=.18, between distribution and step.

Table A.6: Experiment 2A place condition model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.17	0.21	0.8	>0.4
Distribution	0.44	0.20	2.2	**<.05**
Condition	−0.59	0.35	−1.7	>0.1
VOT Step	0.25	0.01	39.2	**<.0001**
Distribution x Condition	4.13	0.68	6.1	**<.0001**
Distribution x Step	0.05	0.01	3.7	**<.001**
Condition x Step	0.03	0.01	2.4	**<.05**
Distribution x Condition x Step	0.19	0.03	6.9	**<.0001**

Note: The maximum correlation between the fixed factors was r=.14, between distribution and step.

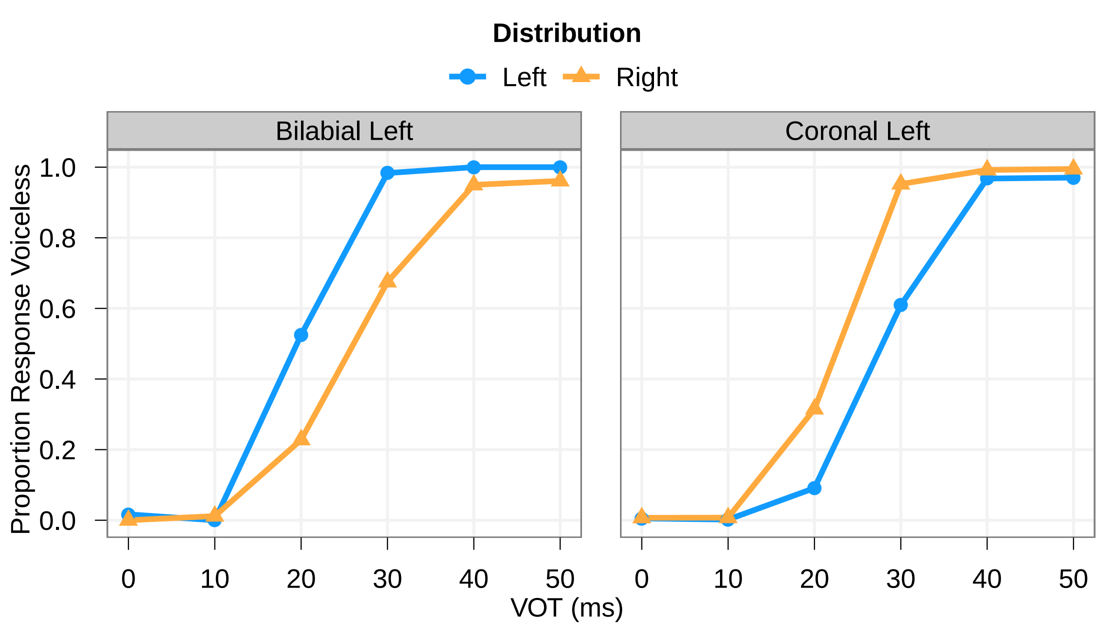

Figure A.2: Experiment 2A distribution by place condition.

Table A.7: Experiment 2A simple effects for place condition.

		Coef β	SE(β)	**z**	**p**
Bilabial Left	Intercept	0.63	0.25	2.5	**<.05**
	Distribution	−2.71	0.44	−6.2	**<.0001**
	VOT Step	0.22	0.01	18.7	**<.0001**
	Distribution x Step	−0.10	0.02	−4.4	**<.0001**
Coronal Left	Intercept	0.02	0.24	0.1	>0.9
	Distribution	1.43	0.34	4.2	**<.0001**
	VOT Step	0.25	0.01	34.4	**<.0001**
	Distribution x Step	0.09	0.02	5.9	**<.0001**

Note: The correlations in the models were r=-.22 for the bilabial left condition and r=.16 for the coronal left condition.

To determine why this was the case we ran an analysis on the effect of place condition (which place was shifted to the left). Bilabial left was coded as 0 and coronal left as 1. We collapsed across day and half to simplify this model (reported in Table A.6). We found a significant effect of distribution (β=.44, p<.03) and an interaction between distribution and condition (β=4.13, p<.0001). The interaction, illustrated in Figure A.2, indicated that the distribution effect was reversed between listeners in the two place conditions: listeners who heard bilabials shifted left showed a boundary difference in the predicted direction, while listeners who heard coronals shifted left showed a reversal. Follow up analyses (Table A.7 confirmed that both these effects were significant (Bilabials: β=-2.71, p<.0001; Coronals: β=1.43, p<.0001).

A.3 Discussion

Experiment 2A aimed to test phoneme-specific learning. While we found that participants did show boundary differences between the continua with different onsets, these did not seem to be due to training as half of the listeners showed boundaries opposite the distributions they were trained on. We thought that listeners might be more successful in learning the boundary shifts if we re-designed the experiment so that they would need to pay attention to place as well as voicing in order to perform the task. This was Experiment 2, described in Chapter 3.

APPENDIX B
EXPERIMENT 4A

The experiment reported in this Appendix was the first attempts to get at the question better answered by Experiment 4, which tested word-specific boundary learning. The initial design (Experiment 4A) had a similar problem to Experiment 2A, in that participants had little reason to pay attention to lexically specific information. Since we had already run and thought about the results of Experiment 2A, we were able to more quickly move to a new version of this experiment, which is why Experiment 4A has so few participants.

Experiment 4A addresses Aim 3: to test whether word-specific perceptual learning is possible. The Introduction and Chapter 3 discuss this possibility and the theoretical implications of this type of learning in more detail. The goal was to expose listeners to some words with a VOT distribution shifted to the left, and other words with distributions shifted to the right. Critically, both sets of words shared the same consonant-vowel (CV) onset. If listeners can learn VOT distributions that apply to individual words, then they should show different category boundaries for words in the two different distributions.

B.1 Method

B.1.1 Design

Experiment 4A participants heard words with the same CV onset shifted in opposite directions. For each participant, half of the continua were heard with VOTs from the left distribution (the left-shift words) and the other half from the right distribution (the right-shift words). Table B.1 shows the experimental and filler words used for the experiment: words with /b/ and /p/ onsets are the experimental items and words with /l/ and /r/ onsets are fillers. These item sets can be divided into three pairs where

Table B.1: Experiment 4A stimulus items

	/b/	/p/	/l/	/r/
	beach	peach	lace	race
left-distribution	bin	pin	lake	rake
	bug	pug	lane	rain
	beak	peak	lei	ray
right-distribution	bill	pill	lock	rock
	buck	puck	lamp	ramp

the experimental words share a CV onset. Each participant was assigned to one of two lists in which one item set from each of the three CV onset pairs occurred in the left-distribution, and the other set with the same onset CV occurred in the right-distribution. One of these lists is shown in Table B.1, and the other list was the same but with the opposite distributions for each item set. (check that the combination of sets listed was one of the lists)

Listeners completed a total of 324 critical trials per day, 162 in each shift direction. Distributions held within each item set instead of only across item set, so each continuum had 54 trials each day. The number of repetitions per continuum at each step is shown in Table B.2, and the number of repetitions for the shift direction as a whole is three times these numbers. Unlike Experiment 1, both distributions extended the full width of the continuum so there were no steps in either distribution that had 0 exemplars. This was done to eliminate the possibility that boundary shift affects would occur because of differences in the range of the continua. In addition to the experiment trials there were an equal number of filler trials per item set, which were split evenly between the /l/ and /r/ filler words for a total of 324 filler trials per day. In total, listeners

Table B.2: Experiment 4A VOT distributions.

	-30	-20	-10	0	10	20	30	40	50	60	70	80
left distribution	1	6	12	6	1	1	6	12	6	1	1	1
right distribution	1	1	1	6	12	6	1	1	6	12	6	1

completed 648 trials per session and participated in two sessions each (1296 trials). The second session was identical to the first except for the order of trials which was randomized for each session.

B.1.2 Participants

Participants were 9 individuals from the University of Iowa community who participated in the study in exchange for course credit or pay–the reason for the small number of participants in this study is discussed later in the discussion section for this experiment. All participants were monolingual native English speakers who reported normal hearing and normal or corrected-to-normal vision. Informed consent was obtained in accordance with University and APA standards. While the majority of participants (8) completed both sessions of the study, 1 participant completed only one of the two sessions and was therefore excluded from analysis.

B.1.3 Stimuli

B.1.3.1 Auditory Stimuli

Auditory stimuli consisted of six twelve-step /b/ to /p/ VOT continua ranging from -30 to 80ms. Two of the continua were used in Experiment 1 and the remaining continua were new. The new continua were created in the same manner as those in Experiment 1 (by cross-splicing recordings of natural speech). The recording and cross-splicing methods used to create the stimuli are described in Chapter 2.

B.1.3.2 Visual Stimuli

Pictures representing each item listed in Table B.1 were constructed using the same picture norming technique described for Experiment 1 in Chapter 2. The final images were approved by the author and thesis supervisor.

B.1.4 Procedure

The procedure was identical to that used in Experiment 1, and is described in Chapter 2. Because of the different number of trials in this experiment, drift correct events (to compensate for small movements of the eye-tracker) occurred every 24 trials instead of every 30. This split the 648 trials per session 27 blocks.

B.2 Results

A boundary shift is predicted such that categorization data for words trained on the left distribution will have a boundary shifted towards the left, while data for the other words (those trained on the right distribution) will show a boundary shifted towards the right.

The structure of the results section mirrors that used for previous experiments, where each section addresses a different question. The same mixed-effects modeling strategy is also employed. The first part of the results section addresses overall task performance and the second section assess perceptual learning of shifted boundaries over the course of the experiment.

B.2.1 Task Performance

Mouse-click responses were first examined to assess overall performance on the task. On experimental trials where the stimulus began with a /b/ or /p/, participants clicked on a filler item on only .14% of the trials, which indicates that they were paying attention to both the auditory and visual stimuli. Performance on the unambiguous

Table B.3: Experiment 4A percent correct at endpoints for each continuum.

continuum	/b/	/p/
beach	99.8	99.7
beak	99.0	99
bill	99.8	99.3
bin	99.9	99.1
buck	98.8	100
bug	99.8	98.7

continuum endpoints was also good. On the /b/ side (-30 to -10 ms), participants were 99% correct. On the /p/ side (60 to 80 ms), participants were 98.7% correct. Endpoint performance was also high when broken down by continuum (Table B.3).

B.2.2 Perceptual Learning

The most critical analysis concerns learning of the shifted distributions. Response (voiced or voiceless, coded as 0 or 1 respectively) was the dependent variable. VOT step, distribution, and day were fixed factors coded and centered as in previous experiments. Participant and continuum were treated as random-effects. The random slopes models failed to converge and adding random intercepts for continua as well improved the model fit (χ^2(2)=21.3, p>.6), so we selected this model.

The perceptual learning model with the best fit to the data is reported in full in Table B.4. The model showed a main effect of distribution (β=-1.61, p<.0001) in the direction predicted by the distribution manipulation. A marginal interaction between distribution, day, and half (β=-1.38, p<.06) suggested that the learning effect for the two halves of each day was not the same on both days. A plot of the data (Figure B.1) revealed that contrary to expectations, the boundary shifts looked reversed or smaller in the sec-

Table B.4: Experiment 4A perceptual learning model.

	Coef β	SE(β)	**z**	**p**
Intercept	0.67	0.28	2.4	**<.05**
Distribution	−1.61	0.19	−8.5	**<.0001**
Day	0.14	0.18	0.8	>0.4
Half	0.06	0.18	0.3	>0.7
VOT Step	0.20	0.01	26.6	**<.0001**
Distribution x Day	−0.48	0.36	−1.3	>0.2
Distribution x Half	0.43	0.36	1.2	>0.2
Day x Half	−0.17	0.36	−0.5	>0.6
Distribution x Step	−0.03	0.01	−2.0	**<.05**
Day x Step	−0.02	0.01	−1.3	>0.2
Half x Step	0.03	0.01	2.4	**<.05**
Distribution x Day x Half	−1.38	0.72	−1.9	>0.1
Distribution x Day x Step	0.01	0.03	0.4	>0.7
Distribution x Half x Step	0.02	0.03	0.8	>0.4
Day x Half x Step	−0.06	0.03	−2.1	**<.05**
Distribution x Day x Half x Step	−0.01	0.06	−0.1	>0.9

Note: The maximum correlation between the fixed factors was r=.046, between distribution and step.

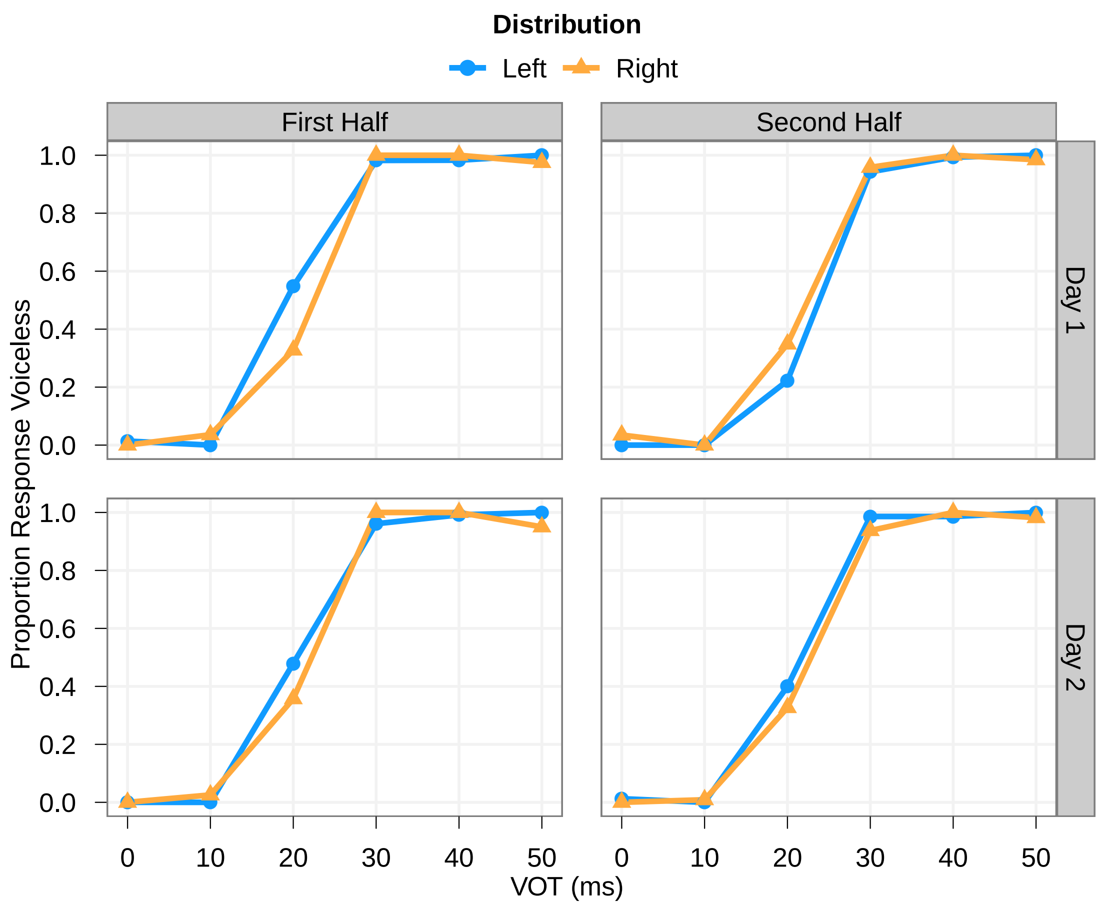

Figure B.1: Experiment 4A distribution effect by experiment day and half.

ond half of each training session. Follow-up analyses (Table B.5) revealed an effect of distribution for both days (Day 1: β=-1.35, p<.0001; Day 2: β=-1.89, p<.0001) and a distribution by half interaction for the first day (β=1.17, p<.03). This interaction indicated that as suggested by the figure, the boundary shift in the second half of the first day was smaller than the shift in the first half, so listeners did not show evidence for perceptual learning of the boundary shifts based on the distributions to which they were exposed.

B.3 Discussion

Experiment 4A aimed to test lexically specific learning. Instead, we failed to find evidence of sensitivity to our distributional manipulation: although there was a difference between the boundaries for words in the left- and right-shifted distributions, this seemed to grow smaller instead of larger over time. We thought that a different design might encourage participants to pay attention to the distributions of each continuum, so we ran a new version of this experiment to address the same question. This was Experiment 4, described in Chapter 4.

Table B.5: Experiment 4A simple effects for days one and two.

		Coef β	SE(β)	**z**	**p**
Day 1	Intercept	0.59	0.27	2.2	**<.05**
	Distribution	−1.35	0.26	−5.1	**<.0001**
	Half	0.14	0.26	0.5	>0.6
	VOT Step	0.21	0.01	17.3	**<.0001**
	Distribution x Half	1.17	0.52	2.3	**<.05**
	Distribution x Step	−0.04	0.02	−1.5	>0.1
	Half x Step	0.07	0.02	2.8	**<.01**
	Distribution x Half x Step	0.02	0.05	0.5	>0.6
Day 2	Intercept	0.74	0.31	2.4	**<.05**
	Distribution	−1.90	0.27	−7.0	**<.0001**
	Half	−0.03	0.25	−0.1	>0.9
	VOT Step	0.19	0.01	21.1	**<.0001**
	Distribution x Half	−0.24	0.51	−0.5	>0.6
	Distribution x Step	−0.02	0.02	−1.4	>0.2
	Half x Step	0.00	0.02	0.3	>0.8
	Distribution x Half x Step	0.02	0.03	0.6	>0.5

Note: The maximum correlations in both models were between distribution and step, and were r=.18 for the first day and r=-.11 for the second day.

REFERENCES

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171–3183.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552.

Apfelbaum, K. S., & McMurray, B. (2011). Using Variability to Guide Dimensional Weighting: Associative Mechanisms in Early Word Learning. *Cognitive Science*, 1–2.

Bates, D. (2005). Fitting Linear Mixed Models in R. *R News*, *5*(1), 27–30.

Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, *13*, 61–81.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Buchholz, L. K. (2009). *Perceptual Learning of Dysarthric Speech: Effects of Familiarization and Feedback*. Unpublished doctoral dissertation, The University of British Columbia.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Clarke, C. M., & Luce, P. A. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. In *Isca workshop on plasticity in speech perception* (pp. 23–26). London.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, *70*(4), 604–618.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Cole, J., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, *38*(2), 167–184.

Connine, C. M. (2004). It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic bulletin & review*, *11*(6), 1084–1089.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics*, *70*(3), 403–411.

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*(1-2), 1–27.

Cutler, A. (2010). Abstraction-based Efficiency in the Lexicon. *Laboratory Phonology*, *1*(2), 301–318.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 91–111). Berlin: de Gruyter.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In *Proceeding of interspeech* (pp. 2056–2056).

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology. Human perception and performance*, *6*(1), 110–125.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

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

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3-4), 305–320.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*(3-4), 423–445.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of experimental psychology. Human perception and performance*, *23*(2), 481–503.

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature*, *408*(6815), 927–928.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3-4), 373–405.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of experimental psychology: Human perception and performance*, *34*(2), 460–474.

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology Human Perception and Performance*, *37*(1), 283–295.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic bulletin & review*, *18*(5), 943–50.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego, CA: Academic Press.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, *51*(2), 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262–268.

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. (2008). First impressions and last resorts: how listeners adjust to speaker variability. *Psychological science*, *19*(4), 332–8.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

Labov, W. (1994). *Principles of linguistic change* (Vol. 1) (No. 1). Cambridge, MA: Blackwell.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word Journal Of The International Linguistic Association*, *20*(3), 384–422.

Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, *26*(4), 381–400.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, *32*(3), 543–562.

Maye, J., & Gerken, L. (2000). Learning Phonemes Without Minimal Pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual boston university conference on language development* (Vol. 2, pp. 522–533). Cascadilla Press.

Maye, J., & Gerken, L. (2001). Learning phonemes: how far can the input take us? In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th annual boston university conference on language development.* (pp. 480–490).

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognit Psychol*, *18*(1), 1–86.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McMurray, B., Cole, J., & Munson, C. (2011). Features as an emergent product of perceptual parsing: Evidence from V-to-V coarticulation. In G. N. Clements & R. Ridouane (Eds.), *Where do features come from?* (pp. 197–235).

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–46.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive science*, *30*(6), 1113–26.

McQueen, J. M., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. In *Isca workshop on plasticity in speech perception* (pp. 233–236). London.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, *13*(6), 958–965.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*(2), 202–240.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, *109*(3), 1181–1196.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Oden, G., & Massaro, D. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172–191.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, *130*(1), 461–472.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency , lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pisoni, D. (1997). Some thoughts on ‚Äúnormalization‚Äù in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.

Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *Journal of experimental psychology: Human perception and performance*, *35*(4), 1205–1220.

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics*, *7*(4), 393–402.

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, *2*(3), 181–194.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211.

Skoruppa, K., & Peperkamp, S. (2011). Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *New York*, *35*(2), 348–366.

Summerfield, Q. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *The Journal of the Acoustical Society of America*, *62*(2), 435–448.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131–136.

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, *75*(3, Pt 1), 983–992.

Syrdal, A. K. (1996). Acoustic variability in spontaneous conversational speech of American English talkers. In *Proceeding of fourth international conference on spoken language processing icslp 96* (pp. 438–441). Ieee.

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *Journal of the Acoustical Society of America, 128*(4), 2090–2099.

Toscano, J. C., & McMurray, B. (2011). Cue integration and context effects in natural and synthetic speech. Manuscript under review.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change, 12*(1), 51–68.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.