Attention and salience in lexically-guided perceptual learning

by

Michael McAuliffe

B.A., University of Washington, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Linguistics)

The University of British Columbia
(Vancouver)

July 2015

Abstract

Psychophysical studies of perceptual learning find that perceivers only improve the accuracy of their perception on stimuli similar to what they were trained on. In contrast, speech perception studies of perceptual learning find generalization to novel contexts when words contain a modified ambiguous sound. This dissertation seeks to resolve the apparent conflict between these findings by framing the results in terms of attentional sets. Attention can be oriented towards comprehension of the speaker's intended meaning or towards perception of a speaker's pronunciation. Attention is proposed to affect perceptual learning as follows. When attention is oriented towards comprehension, more abstract and less context-dependent representations are updated and the perceiver shows generalized perceptual learning, as seen in the speech perception literature. When attention is oriented towards perception, more finely detailed and more context-dependent representations are updated and the perceiver shows less generalized perceptual learning, similar to what is seen in the psychophysics literature. This proposal is supported by three experiments. The first two implement a standard paradigm for perceptual learning in speech perception. In these experiments, promoting a more perception-oriented attentional set causes less generalized perceptual learning. The final experiment uses a novel paradigm where modified sounds are embedded in sentences during exposure. Perceptual learning is found only when the modified sound is embedded in words that are not predictable from the sentence. When modified sounds are in predictable words, no perceptual learning is observed. To account for this lack of perceptual learning, I hypothesize that sounds in predictable sentences are less reliable than sounds in words in isolation or unpredictable sentences. In the cases where perceptual learning is present, contexts which support comprehension-

oriented attentional sets show larger perceptual learning effects than contexts promoting perception-oriented attentional sets. I argue that attentional sets are a key component to the generalization of perceptual learning to new contexts.

Preface

All of the work presented henceforth was conducted in the Speech in Context Laboratory at the University of British Columbia, Point Grey campus. All experiments and associated methods were approved by the University of British Columbia's Research Ethics Board [certificate #H06-04047].

I was the lead investigator for all experiments. Jamie Russell and Jobie Hui aided in data collection for the experiments in Chapter 2. Jobie Hui and Michelle Chan were involved in stimulus preparation and data collection for the experiment in Chapter 3. Molly Babel was involved throughout all experiments in concept formation and manuscript edits.

Table of Contents

List of Tables

List of Figures

Acknowledgments

This dissertation would not have been possible without support from innumerable sources. First and foremost, I must thank Molly Babel, whose enthusiasm and passion for linguistic research is infectious. Equally impressive is her focus and organization, which really helped get this dissertation done in (somewhat) timely fashion. I could not imagine a better supervisor and collaborator. Additionally, feedback and discussions from my committee members, Eric Vatikiotis-Bateson and Carla Hudson Kam, have been invaluable in shaping my thinking on all of my projects over the past five years.

Several seminars provided excellent perspectives on my dissertation research. The first, taught by Martina Wiltschko and Michael Rochemont, helped shape my initial prospectus. The second, taught by Molly Babel and Kathleen Currie Hall, provided a forum to share my intiial findings and thought process and helped refine them. I am also grateful to my fellow students in those seminars for all the lively discussions. My fellow graduate students have shaped my thinking and my graduate experience, namely: Kevin McMullin, Mark Scott, Anita Szakay, Jen Abel, Alexis Black, Heather Bliss, Blake Allen, Michael Fry, Scott Mackie, and Murray Schellenberg. You're all awesome.

I would like to thank the members of the Speech in Context lab, who are uniformly wonderful people. In particular, Michelle Chan and Jobie Hui helped me with stimuli selection, and Jobie Hui and Jamie Russell helped me with running participants. I am indebted to all of you.

My wonderful parents have supported me from the beginning of studying linguistics, even if it wasn't the most viable major for a career in undergrad. I certainly would not have found my passion in linguistics if it weren't for the exposure

to other languages and cultures that I had growing up. My brother, aunts, uncles and cousins from around the world have been amazingly supportive as well. Thank you especially to Laura Tammpere who has been a constant source of support and fun. I love you all!

Chapter 1

Introduction

Listeners are faced with a large degree of phonetic variability when interacting with their fellow language users. Speakers differ in size, gender, and sociolect, which makes speech sound categories overlap in acoustic dimensions. Despite this variation, listeners can interpret disparate and variable productions as belonging to a single word type or sound category, a phenomenon referred to as perceptual constancy (Shankweiler et al., 1977; Kuhl, 1979) or recognition equivalence (Sumner and Kataoka, 2013). One of the processes for achieving this constancy is perceptual learning, whereby perceivers update a perceptual category based on contextual factors.

While perceptual learning is a response to speaker variation, there is also variation on the part of the listener. For instance, a professor with a non-native accent can cause shifts in their students' attention. Some students may be unphased and focus on the content of the lecture. Some may focus on the unfamiliar pronunciations in order to better understand the professor, while others would be distracted by the unfamiliarity. Perceptual learning is typically framed in terms of speaker variation. In contrast, this dissertation examines the effect of listener variation on perceptual learning.

In the speech perception literature, perceptual learning has two distinct, yet related, usages. It can refer to the process of learning to understand a group of speakers that share a common characteristic, such as a nonnative accent (Bradlow and Bent, 2008). The second usage, which this dissertation adopts, is more

constrained. Perceptual learning here refers to a listener updating their perceptual categories following exposure to a single speaker with a modified sound category (Norris et al., 2003).

The primary locus of investigation within the perceptual learning literature is generalization to novel contexts. In most studies, exposure to modified sound categories generalizes to other words and nonwords when the modified exposure tokens are embedded in real words (Norris et al., 2003; Reinisch et al., 2013). This paradigm is referred to as lexically-guided perceptual learning, as listeners are exposed to the modified sounds in the context of real words in a lexical decision task. In a lexical decision task, participants hear words (e.g. *silver* in English) or nonwords (e.g. *shilver*). For each auditory token, they are asked whether they heard a word in English or not. On the other hand, generalization appears to be more limited when perceptual learning is induced through visually-guided paradigms. In visually-guided paradigms, listeners are exposed to a modified category in a nonword (i.e., a token halfway between *aba* and *ada*) matched to an unambiguous video signal (i.e., a person saying *aba*). The perceptual learning exhibited from these visually-guided experiments is found to influence only the specific nonword that perceivers are exposed to and not other similar nonwords (Reinisch et al., 2014). This kind of exposure-specificity effect has been widely reported in perceptual learning studies in the psychophysics literature (Gibson, 1953, for review).

Why, then, does lexically-guided perceptual learning produce such generalization? Here I posit that these results can be understood by considering the attentional set exploited in the exposure phase. An attentional set is a strategy that is employed by perceivers to prioritize certain aspects of stimuli. Attentional sets can be induced through instructions or through properties of stimuli themselves. A common example in the visual domain is the attentional sets employed in visual search tasks. If a perceiver has to find a target shape in a field of shapes, there are two possible attentional sets (Bacon and Egeth, 1994). The first is a singleton-detection attentional set: a diffuse set where any salient information along any dimension will be given priority. If the target shape is a circle in a field of squares, then the singleton-detection attentional set is elicited, because the shape to find is always a highly salient element. Singleton detection can result in slowed reaction times if a distractor singleton (i.e. a red square) is present. The second atten-

tional set is the feature-detection set: a focused set limited to the target's defining feature (i.e. the color red or the square shape). If there are multiple redundant targets or many distractor singletons, then the singleton-detection attentional set will not be an effective strategy and feature detection will be employed. Using feature detection, participants are not distracted by singletons. Bacon and Egeth (1994) speculate that singleton detection is the default attentional set. Feature detection is only employed when singleton detection is made ineffective through stimulus design.

In the speech perception literature, two broad attentional sets have been posited (Cutler et al., 1987; Pitt and Szostak, 2012). The first is a *comprehension-oriented* or *diffuse* attentional set – this is the attentional set assumed to operate during normal language use. When oriented towards comprehension, listeners are focused on comprehending the intended message of the speech, and a comprehension set is promoted by tasks that focus on word identity and word recognition. The comprehension-oriented attentional set is elicited in lexically-guided perceptual learning paradigms through their use of lexical decision tasks and the embedding of modified sound categories in word tokens. A second kind of attentional set is a *perception-oriented* or *focused* attentional set, where a listener is focused more on the low-level signal properties of the speech rather than the message. The perception-oriented attentional set is promoted by tasks such as phoneme/syllable monitoring or mispronunciation detection. The tasks used in visually-guided perceptual learning and perceptual learning within the psychophysics literature can be thought of as eliciting this attentional set. Stimuli used in these paradigms are either nonwords or visual stimuli, and so lack linguistic meaning.

Comprehension and perception are, of course, interconnected concepts. For the purposes of this thesis, I largely follow the distinction drawn in Pitt and Szostak (2012). In this distinction, comprehension refers to processing and understanding a speaker's intended meaning. In other words, it is the activation or retrieval of *linguistic objects* that have been abstracted away from the specific instances. Perception is defined as hearing and processing the speech as pronounced. It is then the encoding of the fine details of a specific instance of an abstract linguistic category. If a listener hears the word *cat* ([kæ̰ˀ]), the listener both perceives the fine details of the specific production (i.e., an unreleased [t], glotalization on the vowel,

gender of the speaker, etc.) and comprehends the lexical item "cat". Perception and comprehension do not always occur together, however. If a listener hears a nonword like *keet* ([kit]), the listener still perceives the details, but there is no lexical item to comprehend. Conversely, listeners can comprehend items that are not actually produced. For instance, in a conversation where one person has a cat, they can say "I have to go home to feed my, uh, you know." The listener can still comprehend the speaker's meaning of feeding their cat, without an actual pronunciation of the word *cat*. In terms of theories of speech perception, comprehension maps to the abstract linguistic representations (sound categories, words, etc.), and perception to the fine-detailed episodic traces. Attention to linguistic properties (i.e. syntactic category) or signal properties (i.e. the speaker's gender) has been shown to change the relative strengths of encoding for abstract or episodic representations (Goldinger, 1996; Theodore et al., 2015). These differences in attention correspond well to the attentional sets proposed above for comprehension and perception.

The core hypothesis of this dissertation is that perceivers who adopt a more comprehension-oriented attentional set will show more generalization than those who adopt a more perception-oriented attentional set. This hypothesis captures the basic findings of perceptual learning in the speech perception and psychophysics literature. Comprehension-oriented exposure tasks (e.g., lexical decision) lead to generalization on novel test items, but perception-oriented exposure tasks (e.g., speech reading) do not generalize to novel test items. To test the hypothesis, I use a lexically-guided perceptual learning paradigm to expose listeners to an /s/ category modified to sound more like /ʃ/. Groups of participants differ in whether comprehension-oriented or perception-oriented attentional sets are favored when processing the modified /s/ category based on experimental manipulations. The favoring of attentional sets are implemented in four ways across the three experiments presented in this dissertation. Two manipulations are linguistic in nature, one is instruction-based, and the fourth is stimulus-based. The rest of this section is devoted to an overview of these manipulations and their motivations.

Before introducing the manipulations themselves, a definition of perceptual salience is necessary. Salience is a widely-used term and is poorly-defined across the literature. For the purposes of this dissertation I adopt the following definition: an element is salient if it is unpredictable given the context and/or easily distin-

guishable from other possible elements. The modified /s/ category that is being learned in this dissertation already has salient signal properties in the form of high frequency, relatively high amplitude, aperiodic noise. Increasing the perceptual salience of the modified /s/ category is a function of embedding it in a linguistic position with little conditioning context or increasing the acoustic distance from a typical /s/ production.

I argue that increased perceptual salience promotes a more perception-oriented attentional set. In general, psycholinguistic studies have a small number of target trials and a large number of unrelated filler trials. By overwhelming the target trials with filler trials, it is assumed that participants will be unlikely to notice of the true purpose of the experiment until debriefing. Changing the nature of the filler trials can induce attentional set changes: if more fillers are words rather than non-words, a comprehension-oriented attentional set is promoted (Mirman et al., 2008). Increasing the number of target trials can also induce attentional set changes. Instructing participants to attend to a modified sound has less of an effect when there are proportionally more of those trials (Pitt and Szostak, 2012). Put another way, if the modified sounds are salient due to prevalence, participants are more likely to notice the targets and adopt a more perception-oriented attentional set without explicit instructions. Increasing perceptual salience through stimulus design in this dissertation is predicted to have the same effect.

The first linguistic manipulation used to promote different attentional sets is the position of the modified /s/ in the exposure words. Accurate perception is most critical when expectations are low, as perception of highly expected elements serves a more confirmatory role (Marslen-Wilson and Welsh, 1978; Gow and Gordon, 1995). Some groups of participants are exposed to the modified sound only at the beginnings of words (e.g. *silver*, *settlement*) and other groups are exposed to the category only in the middle of words (e.g. *carousel*, *fossil*). Word-initial positions lack the expectations afforded to the word-medial positions, and lexical information exhibits less of an effect on word-initial positions as compared to later positions (Pitt and Samuel, 2006). As such, I predict word-initial exposure will promote a more perception-oriented attentional set through increased the perceptual salience of the modified /s/ category within the task. In contrast, I predict word-medial exposure will promote a more comprehension-oriented attentional

set. Further background on manipulating word position is given in Section 1.2.1.

The second linguistic manipulation employed is the context in which a word appears. In Experiments 1 and 2, participants are exposed to the modified sound category in isolated words, as in previous work (Norris et al., 2003). However, in Experiment 3, the words containing the sound category have been embedded in sentences that are either predictive or unpredictive of the target word. Use of sentence frames is predicted to promote comprehension-oriented attentional sets more than words in isolation. Increasing the predictability of a word increases the expectations for the sounds in those words as well, mirroring the word-position manipulation above. Further background on the use of the sentence frames is given in Section 1.2.2.

Participants in all three experiments receive the same general instructions for the exposure task, but one group of participants in each experiment receive additional instructions about the nature of the /s/ category, following previous studies (Pitt and Szostak, 2012). Without any additional instructions, the task is predicted to promote a comprehension-oriented attentional set. The instructions about /s/ are expected to promote a perception-oriented attentional set. Section 1.3 contains background on attention and instructions.

The stimulus-based manipulation is the degree of typicality of the modified /s/ category – Experiments 1 and 2 differ in this respect. In line with previous work (Norris et al., 2003), participants in Experiment 1 are exposed to a modified category halfway between /s/ and /ʃ/. Participants in Experiment 2 are exposed to an even more atypical /s/ – the modified fricative is more /ʃ/-like than /s/-like. Exposure to an atypical category is predicted to promote a more perception-oriented attentional set because its atypicality is predicted to be more perceptually salient.

The structure of the thesis is as follows. This chapter provides an overview of relevant literature on perceptual learning (Section 1.1), linguistic expectations (Section 1.2), attention (Section 1.3), and category typicality (Section 1.4) as they relate to and motivate the three experiments of this dissertation. Chapter 2 details two experiments using a lexically-guided perceptual learning paradigm, each with different conditions for levels of lexical bias and attention. The two experiments differ in the acoustic properties of the exposure tokens, with the first experiment using a slightly atypical /s/ category that is halfway between /s/ and /ʃ/. The sec-

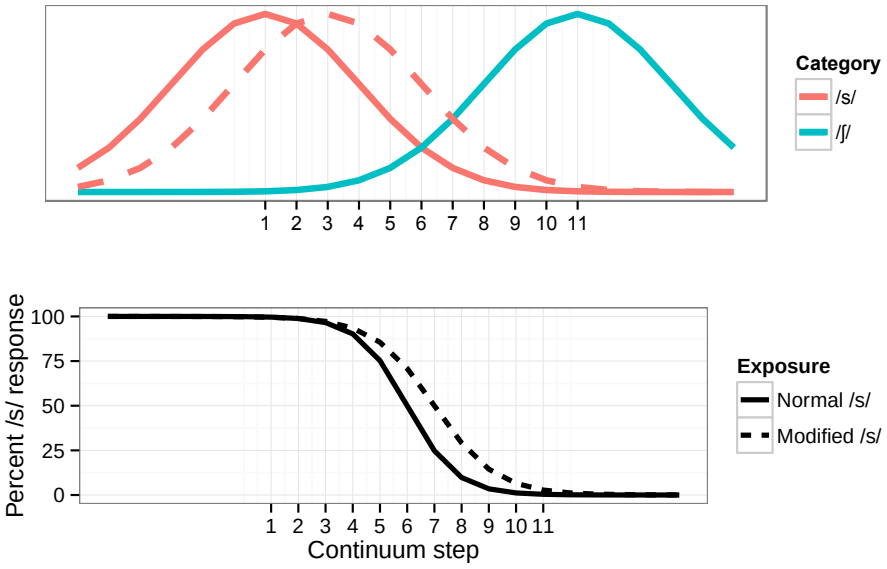
ond experiment uses a more atypical /s/ category that is more /ʃ/-like than /s/-like. Chapter 3 details an experiment using a novel exposure paradigm that manipulates semantic predictability to increase the linguistic expectations during exposure. Finally, Chapter 4 summarizes the results and discusses implications and future directions. The perceptual learning literature has generally used consistent processing conditions to elicit perceptual learning effects, and a goal of this dissertation is to examine the robustness and degree of perceptual learning across conditions that promote more comprehension- or more perception-oriented attentional sets.

1.1 Perceptual learning

Perceptual learning is a well established phenomenon in the psychophysics literature. Training can improve a perceiver's ability to discriminate in many disparate modalities (e.g., visual acuity, somatosensory spatial resolution, weight estimation, and discrimination of hue and acoustic pitch (see Gibson, 1953, for review)). In the psychophysics literature, perceptual learning is an improvement in a perceiver's ability to judge the physical characteristics of objects in the world through attention on the task, but not reinforcement, correction, or reward. Perceptual learning in speech perception refers to updating a listener's sound categories based on exposure to a speaker's modified production category (Norris et al., 2003; Vroomen et al., 2007). Figure 1.1 shows a schema of perceptual learning. Exposure to a speaker exhibiting the modified category's distribution of /s/ (top panel) causes participants to update their perceptual /s/ category to include more /ʃ/-like instances. This expanded category (assuming no modifications to their /ʃ/ category) results in a greater willingness of the participant to categorize ambiguous /s/-/ʃ/ instances as /s/ rather than /ʃ/ (bottom panel). Perceptual learning effects are then evaluated as the difference between the normal categorization function and the one following exposure to a modification.

Norris et al. (2003) began the recent set of investigations into lexically-guided perceptual learning in speech. Norris and colleagues exposed one group of Dutch listeners to a fricative halfway between /s/ and /f/ at the ends of words like *olif* "olive" and *radijs* "radish", while exposing another group to the ambiguous fricative at the ends of nonwords, like *blif* and *blis*. Following exposure, both groups

Figure 1.1: Schema of perceptual learning. The top panel shows categories for /s/ and /ʃ/ along a continuum, with a modified /s/ category in the dashed line. The bottom panel shows a categorization function for exposure to a typical /s/ (solid) and a modified /s/ (dashed).



of listeners were tested on their categorization of a fricative continuum from 100% /s/ to 100% /f/. Listeners exposed to the ambiguous fricative at the end of words shifted their categorization behavior, while those exposed to the same sounds at the end of nonwords did not. The exposure using words was further differentiated by the bias introduced by the words. That is, half the tokens ending in the ambiguous fricative formed a word if the fricative was interpreted as /s/ but not if it was interpreted as /f/, and the others were the reverse. Listeners exposed only to the /s/-biased tokens categorized more of the /f/-/s/ continuum as /s/, and listeners exposed to /f/-biased tokens categorized more of the continuum as /f/. The ambiguous fricative was associated with either /s/ or /f/ according to the bias induced by the word, which led to an expanded category for that fricative at the expense of the other category. These results crucially show that perceptual categories in speech are malleable, and that the lexical system of the listener facilitates generalization to that category in new forms and contexts.

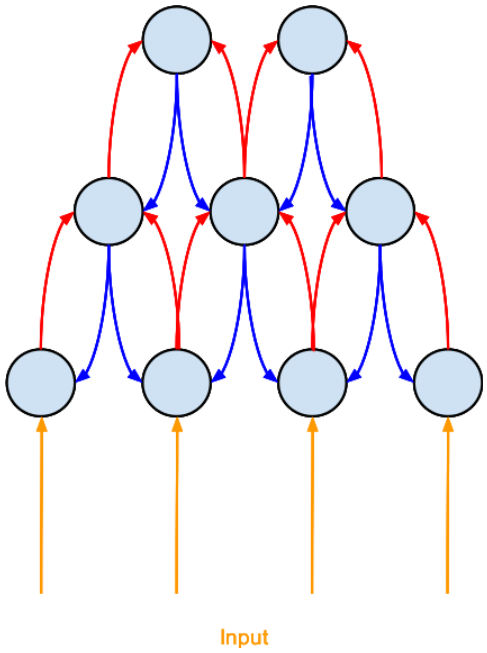
In addition to lexically-guided perceptual learning, unambiguous visual cues to sound identity can cause perceptual learning as well; this is referred to as perceptual recalibration. In Bertelson et al. (2003), an auditory continuum from /aba/ to /ada/ was synthesized and paired with a video of a speaker producing /aba/ or /ada/. Participants first completed a pretest that identified the maximally ambiguous step of the /aba/-/ada/ auditory continuum. In eight blocks, participants were randomly exposed to the ambiguous auditory token paired with video for /aba/ or /ada/. Following each block, they completed a short categorization test. Participants showed perceptual learning effects, such that they were more likely to respond with /aba/ if they had been exposed to the video of /aba/ paired with the ambiguous token in the preceding block, and vice versa for /ada/.

Visually-guided perceptual learning in speech perception has been modeled using a Bayesian belief updating framework (Kleinschmidt and Jaeger, 2011). In this framework, the model categorizes the incoming stimuli based on an acoustic-phonetic feature and a binary visual feature, and then updates the distribution to reflect that categorization. This updated conditional distribution is then used for future categorizations in an iterative process. Kleinschmidt and Jaeger (2011) effectively model the results of the behavioral study in Vroomen et al. (2007) in a Bayesian framework, with models fit to each participant capturing the perceptual recalibration and selective adaptation shown over the course of the experiment. The Bayesian belief updating framework has only been applied to the visually-guided perceptual learning paradigm.

A similar, but more general Bayesian framework for perception and action in cognition is the predictive coding model (Clark, 2013), schematized in Figure 1.2. This framework uses a hierarchical generative model that aims to minimize prediction error between bottom-up sensory inputs and top-down expectations. Mismatches between the top-down expectations and the bottom-up signals generate error signals that are used to modify future expectations. Perceptual learning is the result of modifying expectations to match learned input and reduce future error signals. The lowest levels of the hierarchical model have the most detailed representations. Representations lose detail and become more abstract the higher in the hierarchy they are.

The predictive coding framework is adopted for this thesis rather than theo-

Figure 1.2: Schema of the predictive coding framework adapted from Clark (2013). Representations are hierarchical and are more abstract the higher they are in the hierarchy. Blue arrows represent expectations, red arrows are error signals, and yellow is the actual sensory input.



retical models of speech perception because representations in predictive coding can be more general than linguistic objects. Models of speech perception – understandably – focus primarily on the linguistic representations. Representations are usually proposed to encode both abstract and episodic information (e.g. McLennan et al., 2003). Abstract and episodic information would map to higher and lower levels of a representation in the predictive coding framework, respectively. However, a representation in the predictive coding framework is not limited to the linguistic domain. For instance, individual speakers can be thought of as part of more abstract accent representations. If a listener is exposed to multiple speakers of an accent, they are better at understanding novel speakers of that accent than a listener exposed to just one speaker of that accent (Bradlow and Bent, 2008). Predictions

for sensory input when listening to the novel speaker would then be shaped by expectations based on the abstract accent as well as the linguistic representations generally assumed.

Perceptual learning in the psychophysics literature has shown a large degree of exposure-specificity, where observers show learning effects only on the same or very similar stimuli as those they were trained on. As such, perceptual learning has been argued to reside in or affect the early sensory pathways, where stimuli are represented with the greatest detail (Gilbert et al., 2001). Visually-guided perceptual learning has also shown a large degree of exposure-specificity, where participants do not generalize cues across speech sounds (Reinisch et al., 2014) or across speakers unless the sounds are sufficiently similar (Eisner and McQueen, 2005; Kraljic and Samuel, 2005, 2007; Reinisch and Holt, 2013). Crucially, lexically-guided perceptual learning in speech has shown a greater degree of generalization than what would be expected from a purely psychophysical standpoint. The testing stimuli are in many ways quite different from the exposure stimuli. Participants are trained on multisyllabic words ending in an ambiguous sound, but tested on monosyllabic words (Reinisch et al., 2013) and nonwords (Norris et al., 2003; Kraljic and Samuel, 2005). In these cases, generalization is robust; however, some exposure-specificity has been found when exposure and testing use different positional allophones (Mitterer et al., 2013).

Why is lexically-guided perceptual learning more context-general? The experiments performed in this dissertation provide evidence that this context-generality is the result of a listener's attentional set, which can be influenced by linguistic, instruction, or stimulus properties. A comprehension-oriented attentional set, where a listener's goal is to understand the meaning of speech, promotes generalization and leads to greater perceptual learning. A purely perception-oriented attentional set, where a listener's goal is to perceive specific qualities of a signal, does not promote generalization. The attentional set promoted in the experiments in this dissertation is comprehension-oriented, as the tasks are lexically guided. However, perception-oriented attentional sets will be promoted in some conditions. I predict that perceptual learning will be present across all conditions, but participants in conditions that promote perception-oriented attentional sets should show smaller perceptual learning effects. In terms of a Bayesian framework with error propaga-

tion, a more perception-oriented attentional set would keep error propagation more local, resulting in the exposure-specificity seen more in the psychophysics literature and the visually-guided paradigm. A more comprehension-oriented attentional set would propagate errors farther upward to more abstract representations. In both cases, errors would propagate to where attention is focused, but more abstract representations would be more applicable to novel contexts, leading to the observed context-general perceptual learning. These attentional sets will be explored in more detail in Section 1.3 following an examination of the linguistic factors that will be manipulated in the experiments in Chapters 2 and 3.

1.2 Linguistic expectations and perceptual learning

The linguistic manipulations used to induce different attentional sets are lexical bias and semantic predictability. Chapter 2 presents two experiments using a standard lexically-guided perceptual learning paradigm, which uses lexical bias as the means to link an ambiguous sound to an unambiguous category. In Chapter 3, a novel paradigm is used to further promote use of comprehension-oriented attentional sets. This paradigm embeds words in sentences that differ in their semantic predictability.

1.2.1 Lexical bias

Lexical bias is the primary way through which perceptual learning is induced in the experimental speech perception literature. Lexical bias, also known as the Ganong Effect, refers to the tendency for listeners to interpret a speaker's (noncanonical) production as a meaningful word rather than a nonsense word. For instance, given a continuum from a word to a nonword that differs only in the initial sound (e.g., *task* to *dask*), listeners are more likely to interpret any step along the continuum as the word endpoint rather than the nonword endpoint as compared to a continuum where neither endpoint is a word (Ganong, 1980). This bias is exploited in perceptual learning studies to allow for noncanonical, ambiguous productions of a sound to be readily linked to pre-existing sound categories. In terms of the attentional sets proposed in this dissertation, lexical effects, including lexical bias, arise due to comprehension-oriented attentional sets.

Comprehension-oriented attentional sets are not limited to just comprehension tasks. Lexical effects have also been found for reaction time in phoneme detection tasks: sounds are detected faster in words than in nonwords. However, such lexical effects are dependent on the attentional set being employed. If the stimuli are sufficiently repetitive (e.g. all having the same CV shape) the lexical bias effects disappear (Cutler et al., 1987). The monotony or variation of filler items is sufficient to bias listeners towards perception-oriented or comprehension-oriented attentional sets, respectively. The lexical status of the filler items contributes to attentional set adoption as well. Lexical effects are found when the proportion of word fillers is high, but disappear when the proportion of nonword fillers is high (Mirman et al., 2008).

The lexical effect of interest in this dissertation is lexical bias. The degree to which a word biases the perception of a sound is primarily determined by properties of the word. Longer words show stronger lexical bias than shorter words (Pitt and Samuel, 2006). Continua formed using trisyllabic words, such as *establish* and *malpractice*, were found to show consistently larger lexical bias effects than monosyllabic words, such as *kiss* and *fish*. Pitt and Samuel (2006) also found that lexical bias from trisyllabic words was robust across experimental conditions (e.g., compressing the durations by up to 30%), but lexical bias from monosyllabic words was more fragile and condition dependent. The lexical bias effects shown by monosyllabic words only approached those of trisyllabic words when the participants were told to keep response times within a certain margin and were given feedback when the response time fell outside the desired range. The reaction time monitoring could have added a greater cognitive load for participants, which has been shown to increase lexical bias effects (Mattys and Wiget, 2011). Pitt and Samuel (2006) argue that longer words exert stronger lexical bias from more conditioning information present in longer words, as well as greater lexical competition for shorter words.

Within a given word, different positions have stronger or weaker lexical bias effects. Pitt and Szostak (2012) used a lexical decision task with a continuum of fricatives from /s/ to /ʃ/ embedded in words that differed in the position of the sibilant. They found that ambiguous fricatives later in the word, such as *establish* or *embarrass*, show greater lexical bias effects than the same ambiguous fricatives

embedded earlier in the word, such as *serenade* or *chandelier*. Pitt and Samuel (1993) found that for monosyllabic words, token-final targets produce more robust lexical bias effects than token-initial targets. Lexical bias is strengthened over the course of the word. As a listener hears more evidence for a particular word, their expectations for hearing the rest of that word increase.

One final research paradigm that has investigated lexical biases is phoneme restoration tasks (Samuel, 1981). In this paradigm, listeners hear words with noise added to or replacing sounds and are asked to identify whether noise completely replaced part of the speech or noise was simply added to the speech. Lower sensitivity to noise addition versus noise replacement and increased bias for responding that noise has been added is indicative of phoneme restoration – that is, listeners are perceiving sounds not physically present in the signal. Samuel (1981) identified several factors that increase the likelihood of the phoneme restoration effect. In the lexical domain, words are more likely than nonwords to have phoneme restorations. More frequent words are also more likely to exhibit phoneme restoration effects, and longer words also show greater phoneme restoration effects. The position of the sound in the word also influences listeners' decisions, with non-initial positions showing greater effects. The other influences on phoneme restoration discussed in Samuel (1981), namely the signal properties and sentential context are discussed in subsequent sections.

Lexical bias effects are found across a wide range of tasks involving speech. However, lexical bias effects are not uniform across the word. Various models of lexical access give a large role to the initial sounds in the word (Marslen-Wilson and Welsh, 1978; Gow and Gordon, 1995). In such models, initial perception of sounds plays a disproportionate role in providing lexical identity, allowing later sounds to be perceived in reference to the initially perceived lexical identity. In spontaneous speech, sounds are more likely to be produced canonically in early positions than in later positions (Pitt and Szostak, 2012). In terms of the predictive coding framework (Clark, 2013), decreased lexical bias would result from the lack of (or decreased) expectations from higher levels of representation.

Lexical effects are generally argued to be the result of attentional sets rather than the cause (Cutler et al., 1987; Pitt and Szostak, 2012). I propose that ambiguous sounds that are salient, in this case due to their position in the word, cause

listeners to adopt a more perception-oriented attentional set over the course of the experiment. Under this proposal, increasing the perceptual salience of a few ambiguous sounds is functionally equivalent to having many ambiguous sounds that are not as salient. That is, having some number of modified /s/ tokens at the beginnings of words is similar to having a larger number of modified /s/ tokens at the ends of words (with the same overall number of trials) or modified /s/ tokens that are less typical of /s/. In both cases, the likelihood of the participant noticing the ambiguous sound is higher, and so too is their likelihood of adopting a more perception-oriented attentional set for completing the task. The experiments in this dissertation do not fully test this proposal, as the number of exposure tokens is never manipulated. However, it does predict that less typical modified sounds (discussed in Section 1.4) embedded later in words should produce comparable perceptual learning effects as more typical modified sounds at the beginnings of words. It is important to note that perceptual learning is predicted to occur regardless of exposure location, following previous research (Norris et al., 2003; Kraljic and Samuel, 2005; Kraljic et al., 2008a,b; Clare, 2014), but the degree of perceptual learning is predicted to be less when exposure is at the beginnings of words. Table 1.1 lists the predicted perceptual learning effects for the linguistic manipulations and instruction. Experiments 1 and 2 in Chapter 2 test the predictions related to lexical bias.

Table 1.1: Summary of predictions for size of perceptual learning effects under different linguistic expectations and attention.

	Lexical bias		Semantic predictability	
	Word-initial	Word-medial	Unpredictable	Predictable
Regular attention	Smaller effect	Larger effect	Larger effect	Largest effect
Attention to /s/	Smaller effect	Smaller effect	Smaller effect	Smaller effect

1.2.2 Semantic predictability

In addition to lexical bias, semantic predictability (Kalikow et al., 1977) can increase the predictability of modified sound categories by increasing the predictability of the word containing them. Sentences are semantically predictable when the

words prior to the final word point almost definitively to the identity of that final word. For instance, the sentence fragment *The cow gave birth to the...* from Kalikow et al. (1977) is almost guaranteed to be completed with the word *calf*. On the other hand, a fragment like *She is glad Jane called about the...* is far from having a guaranteed completion beyond being a noun. Despite its name, semantic predictability does not incorporate formal semantic theory, but refers to world knowledge that language users have.

Words that are predictable from context are temporally and spectrally reduced compared to words that are less predictable (Scarborough, 2010; Clopper and Pierrehumbert, 2008). Despite this acoustic reduction, highly predictable sentences are generally more intelligible. Sentences that form a semantically predictable whole have higher word identification rates across varying signal-to-noise ratios (Kalikow et al., 1977) in both children and adults (Fallon et al., 2002), and across native monolingual and early (but not late) bilingual listeners listeners (Mayo et al., 1997). Highly predictable sentences are more intelligible to native listeners in noise, even when signal enhancements are not made, but nonnative listeners require both signal enhancements and high predictability to see any benefit (Bradlow and Alexander, 2007). However, when words at the ends of predictive sentences are excised from their context, they tend to be less intelligible than words excised from non-predictive contexts (Lieberman, 1963).

Semantic predictability has similar effects to lexical bias on phoneme categorization (Connine, 1987; Borsky et al., 1998). In those studies, a continuum from one word to another, such as *coat* to *goat*, is embedded in a sentence frame that semantically coheres with one of the endpoints. The category boundary shifts based on the sentence frame. If the sentence frame cues the voiced stop, more of the continuum is subsequently categorized as the voiced stop and vice versa for the voiceless stop.

In the phoneme restoration paradigm, higher semantic predictability has been found to bias listeners toward perceptually restoring a sound (Samuel, 1981). This increased bias towards interpreting the stimulus as an intact word was also coupled with an increase in sensitivity between the two types of stimuli (i.e. noise added to speech, speech replaced with noise), which Samuel (1981) suggests is the result of a lower cognitive load in predictable contexts. Later work has suggested that in

cases of lower cognitive load, finer phonetic details are encoded (see also Mattys and Wiget, 2011).

To summarize, the literature on semantic predictability has shown largely similar effects as lexical bias in terms of how sounds are categorized and restored. From this, I hypothesize that increasing the expectations for a word through semantic predictability will promote a comprehension-oriented attentional set, as perception of the modified sound category will not be strictly necessary for comprehension. Listeners who are exposed to an /s/ category that is more /ʃ/-like only in words that are highly predictable from context are therefore predicted to show larger perceptual learning effects than listeners exposed to the same category only in words that are unpredictable from context. However, there may be an upper limit for listener expectations when both semantic predictability and lexical bias are high, as committing too much to a particular expectation could lead to garden path phenomena (Levy, 2008) or other misunderstandings. Table 1.1 lists the predicted perceptual learning effects for the linguistic manipulations and instruction. The effect of semantic predictability on perceptual learning is explicitly tested in Chapter 3.

1.3 Attention and perceptual learning

Attention is a large topic of research in its own right, and this section only reviews literature that is directly relevant to perceptual learning and this thesis. Attention has been found to have a role in perceptual learning in the psychophysics literature. Indeed, Gibson (1953) identifies attention as the sole prerequisite to perceptual learning. There is some evidence of short-lived perceptual learning without explicit attention (Watanabe et al., 2001), but the effects are not as robust as for attended perceptual learning. Perceptual learning is not alone in requiring attention; learning statistical regularities in a speech stream likewise depends on auditory attention, either explicitly or through passive listening (Toro et al., 2005; Saffran et al., 1997, but see Finn et al., 2014). The model of attention used in this dissertation is that of attentional sets.

Attentional sets refer to the strategies that the perceiver uses to perform a task. The attentional sets widely used in the visual perception literature do not align completely with the notion of perception-oriented and comprehension-oriented at-

tentional sets used here. For instance, in a visual search task, attending to color, orientation, motion, and size are the predominant strategies (Wolfe and Horowitz, 2004). However, some parallels are present. The two broad categories in the visual perception literatures are *focused* and *diffuse* attentional sets. Focused sets direct attention to components of the sensory input, perceiving the trees instead of the forest. Diffuse sets direct attention to global properties of the sensory input, perceiving the forest instead of the trees. The two attentional sets employed in visual search paradigms introduced above – singleton-detection and feature-detection attentional sets – are diffuse and focused, respectively. Perception-oriented and comprehension-oriented attentional sets have also been referred to as focused and diffuse attentional sets in recent speech perception work. In Pitt and Szostak (2012), a diffuse attentional set is employed when detecting words from nonwords in a lexical decision task, and a focused attentional set is employed when participants' attention is directed to a potentially misleading sound. Attentional set selection is primarily affected by the instructions and the stimuli for the task and they tend to become entrenched over time (Leber and Egeth, 2006).

Listeners can employ different attentional sets depending on the nature of the task, as well as other processing considerations. For instance, listeners can attend to particular syllables or sounds in syllable- or phoneme-monitoring tasks (Norris and Cutler, 1988, and others), and even particular linguistically relevant positions (Pitt and Samuel, 1990). However, even in these low-level, signal based tasks, lexical properties of the signal can exhibit some influence if the stimuli are not monotonous enough to disengage comprehension (Cutler et al., 1987). Additionally, when performing a phoneme categorization task under higher cognitive load, such as performing a more difficult concurrent task, listeners show increased lexical bias effects (Mattys and Wiget, 2011). Stimulus variation in general seems to lead towards a more diffuse, comprehension-oriented attentional set, where the goal is firmly more comprehension-based than low-level perception-based. In comprehension-oriented tasks, such as a lexical decision tasks, explicit instructions can promote a more perception-oriented attentional set. When listeners are told that the speaker's /s/ is ambiguous and to listen carefully to ensure correct responses, they are less tolerant of noncanonical productions across all positions in the word (Pitt and Szostak, 2012). That is, listeners whose attention is directed to

the speaker's sibilants are less likely to accept the modified production as a word than listeners given no particular instructions about the sibilants. While the primary task has a large influence on the type of attentional set adopted, other instructions and aspects about the stimuli can shift the listener's attentional set toward another one.

Attentional sets have been found to affect what aspects of stimuli are perceptually learned in the visual domain. Ahissar and Hochstein (1993) found that, in general, attending to *global* features for detection (i.e., discriminating different orientations of arrays of lines) does not make participants better at using *local* features for detection (i.e., detection of a singleton that differs in angle in the same arrays of lines), and vice versa. Perceptual learning in the visual domain is limited to the aspects of the stimuli to which participants were attending.

Attentional sets have not been directly manipulated in previous lexically-guided perceptual learning literature, but some work has been done on how individual differences in attention control can impact perceptual learning. Scharenborg et al. (2014) presents a perceptual learning study of older Dutch listeners in the model of Norris et al. (2003). In addition to the exposure and test phases, these older listeners completed tests for high-frequeny hearing loss, selective attention, and attention-switching control. Selective attention refers to the ability of the participants to focus on one element in visual string to the exclusion of other (potentially distracting) elements, measured using the Flanker Test (Eriksen and Eriksen, 1974). Attention-switching control is measured using the Trail-Making Test (Reitan, 1958), where participants complete two tasks of connecting dots. In the first task, dots are numbered from 1 to 25 and the trail must go from 1 to 25 in order. In the second task, dots are labeled with either letters or numbers, and the trail must alternate between the two in ascending order (1-A-2-B, etc.). Differences in the time for completion of these two tasks is indicative a participant's attention-switching control, with faster performance in the second task indicative of better attention-switching control.

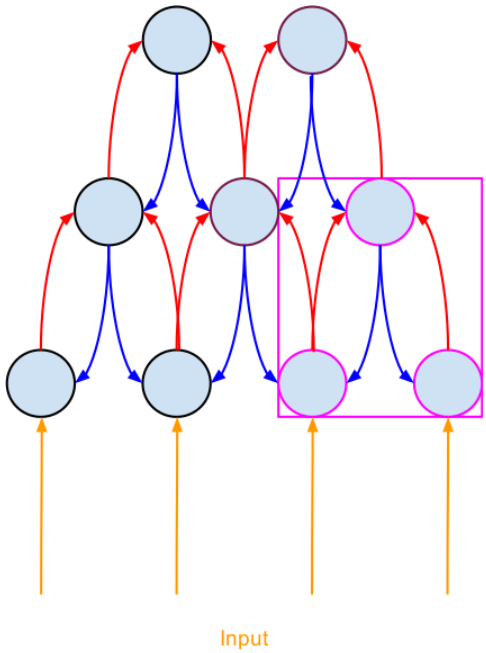
Scharenborg and colleagues found no evidence that perceptual learning was influenced by listeners' hearing loss or selective attention abilities, but they did find a significant relationship between a listener's attention-switching control and their perceptual learning. Listeners with worse attention-switching control showed

greater perceptual learning effects, which the authors ascribed to an increased reliance on lexical information. Older listeners were shown to have smaller perceptual learning effects compared to younger listeners, but the differences were most prominent directly following exposure (Scharenborg and Janse, 2013). Younger listeners initially had a larger perceptual learning effect in the first block of testing, but the effect lessened over the subsequent blocks. Older listeners showed more consistent, but smaller perceptual learning effects, hypothesized to be due to greater prior experience. Scharenborg and Janse (2013) also found that participants who endorsed more of the target items as words in the exposure phase showed significantly larger perceptual learning effects in the testing phase.

There is evidence that attentional sets in the visual domain become entrenched over time (Leber and Egeth, 2006). However, the fact that attention-switching control in older adults was a significant predictor of the size of perceptual learning effects (Scharenborg et al., 2014) reinforces that comprehension and perception, as defined in this dissertation, are not mutually exclusive. These findings do suggest that attention can be switched between comprehension and perception, and that this switching has consequences for perceptual learning. The lexical decision task is oriented towards comprehension, so the primary attentional set is likely to be a diffuse one relying more on lexical information than acoustic. Participants with worse attention-switching control would have been less able to attend to the fine details of the signal than those with better attention-switching control, and it is precisely those with worse attention-switching control that showed the larger perceptual learning effects. The ability to attend to finer sensory representations could prevent error propagation to more abstract representations, leading to a smaller perceptual learning effect for participants with better attention-switching control.

As stated above, Bayesian models account well for the results of perceptual learning experiments. Attentional sets are crucial to the hypothesis tested in this dissertation, but they do not play a role in the conceptual and computational models of perceptual learning. The predictive coding framework (Clark, 2013) provides a gain-based attentional mechanism. Gain is typically likened to increasing the volume. For instance, attending to a specific location on a screen has subjectively similar effects as increasing contrast (Ling and Carrasco, 2006) and attending to speech from a single ear is subjectively similar to increasing the volume for that

Figure 1.3: A schema for predictive coding under a perception-oriented attentional set. Attention is represented by the pink box, where gain is enhanced for detection, but error signal propagation is limited to lower levels of sensory representation where the expectations must be updated. This is represented by the lack of pink nodes outside the attention box. As before, blue errors represent expectations, red arrows represent error signals, and yellow represents the sensory input.



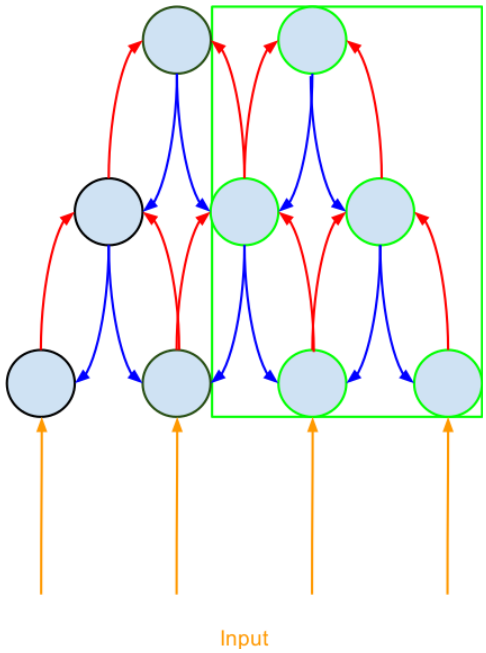
channel. In this model, attention causes greater weight to be attached to error signals from mismatched expectations and sensory input, increasing their weight and their effect on future expectations. However, as noted by Block and Siegel (2013), this view of attention does not capture the full range of experimental results. For instance, in a texture segregation task, spatial attention to the periphery improves detection accuracy where spatial resolution is poor, but attention to central locations, where spatial resolution is high, actually harms accuracy (Yeshurun and Carrasco, 1998). This detrimental effect is an instance of missing the forest for the trees, as spatial resolution increased too much in the central locations to

perceive the larger texture. Instead the fine details interferred with perceiving the larger texture. The attentional mechanism proposed in this dissertation limits error propagation beyond where attention is focused. Attending to perception rather than comprehension should only update expectations about perception of that individual instance. The lower sensory levels are where stimuli are represented with the greatest degree of detail (Gilbert et al., 2001). Perceptual learning at these lower levels should be more exposure-specific and less generalized than any learning that propagates to higher representational levels. Figures 1.3 and 1.4 show schemas of the proposed attention mechanism for updating future expectations under perception-oriented and comprehension-oriented attentional sets, respectively. In contrast, according to the mechanism proposed in Clark (2013), any increases in attention, perception-oriented or otherwise, are predicted to lead to greater perceptual learning.

1.4 Category typicality and perceptual learning

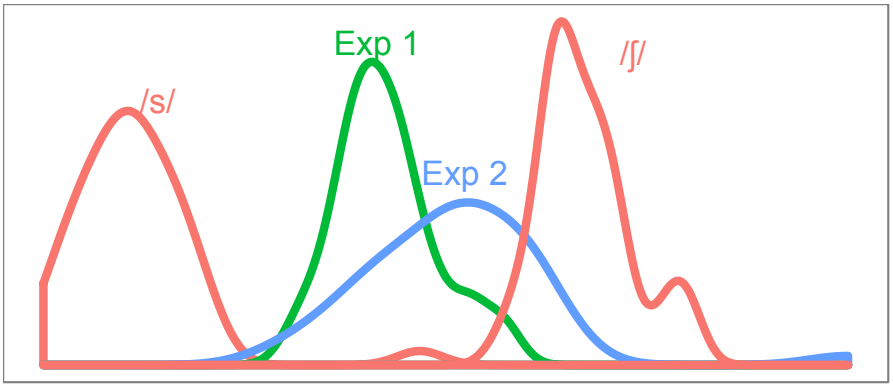
A primary finding across the perceptual learning literature is that learning effects are found only on testing items that are similar in some sense to the exposure items. In the most extreme instance, perceptual learning is only found on the exact same items as exposure (Reinisch et al., 2014), but most commonly, perceptual learning is limited to items produced by the same speaker as the exposure items (Norris et al., 2003; Reinisch et al., 2013). However, a less studied question is what properties of the exposure items cause different degrees of perceptual learning. In this dissertation, two levels of category typicality are used. Figure 1.5 shows four categories. At the left and right ends are the original categories for /s/ and /ʃ/ as produced by a male Vancouver English speaker. The two categories in the middle are the modified categories used in Experiments 1 and 2. Experiment 1 uses a modified /s/ category that is halfway between the original /s/ and /ʃ/, while the modified category for Experiment 2 is skewed more towards /ʃ/. Thus, Experiment 2 uses a more atypical /s/ category (farther from the typical /s/ distribution) than Experiment 1. The more atypical category is predicted to be more salient, and therefore promote a perception-oriented attentional set.

Figure 1.4: A schema for predictive coding under a comprehension-oriented attentional set. Attention is represented by the green box, where it is oriented to higher, more abstract levels of sensory representation. Error signals are able to propagate farther and update more than just the fine grained low level sensory representations. As before, blue arrows represent expectations, red arrows represent error signals, and yellow represents the sensory input.



Sumner (2011) investigated category typicality through a manipulation of presentation order. Listeners were exposed to French-accented English with modifications to the /b/-/p/ category boundary. Participants were exposed to stimuli ranging from English-like to French-like voice onset time for /b/ and /p/. In one presentation order, the order of stimuli was random, but in the others the voice onset time changed in a consistent manner, such as starting as more French-like and becoming more English-like. The presentation order that showed the greatest perceptual learning effects was the one that began more English-like and ended more French-like. The condition that mirrored the more normal course of nonnative speaker pro-

Figure 1.5: Distribution of original categories (red endpoints) and modified /s/ categories used in Experiments 1 and 2. Experiment 1 uses a maximally ambiguous category between /s/ and /ʃ/. Experiment 2 uses a category that is more like /ʃ/ than /s/. The x-axis was generated using acoustic similarities used to generate Figures 2.3 and 2.7.



nunciation changes, starting as more French-like and ending as more English-like, did not produce significantly different behavior than control participants who only completed the categorization task. The random presentation order had perceptual learning effects in between the two ordered conditions. These results suggest that listeners constantly update their category following each successive input, rather than only relying on initial impressions (contra Kraljic et al., 2008b). This finding is mirrored in Vroomen et al. (2007), where participants initially expand their category in response to a single, repeated modified input, but then entrench that category as subsequent input is the persistently same. The data in Vroomen et al. (2007) is modeled using a Bayesian framework with constant updating of beliefs. However, in Sumner (2011), the constantly shifting condition also shows more perceptual learning than a random order of the same stimuli, suggesting that small differences in expectations and observed input induce greater updating than large differences. The bias towards small differences is better captured by the exemplar model proposed by Pierrehumbert (2001), where only input similar to the learned

distribution is used for updating that distribution.

Variability is a fundamental property of the speech signal, so sound categories must have some variance associated with them and certain contexts can have increased degrees of variability. For example, Kraljic et al. (2008a) exposed participants to ambiguous sibilants between /s/ and /ʃ/ in two different contexts. In one, the ambiguous sibilants were intervocalic, and in the other they occurred as part of a /stɹ/ cluster in English words. Participants exposed to the ambiguous sound intervocalically showed a perceptual learning effect, while those exposed to the sibilants in /stɹ/ environments did not. The sibilant in /stɹ/ often surfaces closer to [ʃ] in many varieties of English, due to coarticulatory effects from the other consonants in the cluster (Baker et al., 2011). They argue that the interpretation of the ambiguous sound is done in context of the surrounding sounds, and only when the pronunciation variant is unexplainable from context is the variant learned and attributed to the speaker (see also Kraljic et al., 2008b). In other words, a more /ʃ/-like /s/ category is typical in the context of the /stɹ/ clusters, but is atypical in intervocalic position. Interestingly, given the lack of learning present in in the /stɹ/ context, some degree of salience seems to be required to trigger perceptual learning.

Similarity of input to known distributions has effects in many psycholinguistic paradigms. For instance in phoneme restoration, Samuel (1981) found that the likelihood of restoring a sound increases when said sound is acoustically similar to the noise replacing it. When the replacement noise is white noise, fricatives and stops are more likely to be restored than vowels and liquids. Acoustic signals that better match expectations are less likely to be noticed as atypical.

In this dissertation, the degree of typicality of the modified category is manipulated across Experiments 1 and 2. In one case, the /s/ category for the speaker is maximally ambiguous between /s/ and /ʃ/, but in the other, the category is more like /ʃ/ than /s/. The maximally ambiguous category is hypothesized to be less salient than the more /ʃ/-like /s/ category. This lessened salience will result in greater use of comprehension-oriented attentional sets. I hypothesize that the more /ʃ/-like category will shift listeners' attentional sets to be more perception-oriented due to their greater atypicality, which will lead to less generalized perceptual learning. The predictions for this hypothesis are summarized in Table 1.2.

Table 1.2: Summary of predictions for size of perceptual learning effects when exposed to different typicalities of the modified category.

	Lexical bias			
	Word-initial		Word-medial	
	Less atypical	More atypical	Less atypical	More atypical
Regular attention	Smaller effect	Smaller effect	Larger effect	Smaller effect
Attention to /s/	Smaller effect	Smaller effect	Smaller effect	Smaller effect

1.5 Current contribution

Lexically-guided perceptual learning generalizes to new forms and contexts far more than would be expected from a purely psychophysical perspective (Norris et al., 2003; Gilbert et al., 2001). Lexically-guided paradigms provide a focus on comprehension and psychophysics tasks giving focus to perception, promoting the respecitive attentional sets. Indeed, visually-guided perceptual learning, with its emphasis on perception of speech, shows largely similar exposure-specificity effects as the psychophysics findings (Reinisch et al., 2014). This dissertation expands on the existing literature by modifying the exposure tasks to promote comprehension- or perception-oriented attentional sets. Perceptual learning effects are hypothesized to be smaller in the conditions that promote perception-oriented attentional sets, as perception exposure tasks have shown greater exposure-specificity effects than comprehension exposure tasks.

Chapter 2

Lexical decision

2.1 Motivation

The experiments in this chapter implement a standard lexically-guided perceptual learning experiment with exposure to a modified /s/ category during a lexical decision task. Because the exposure task is one of word recognition, participants are predicted to default to a comprehension-oriented attentional set. Recall that comprehension-oriented attentional sets are hypothesized to facilitate perceptual learning and generalization. Two experimental manipulations guide listeners to use more of a perception-oriented attentional set. The first manipulation relates to the position of the modified /s/ category in the exposure tokens (*silver* versus *carousel*). Lexical bias effects increase as the length of the word increases and as the word unfolds (Pitt and Samuel, 2006; Pitt and Szostak, 2012), so we predict that more learning will take place in *carousel*-like words. The second manipulation is through explicit instructions about the modified /s/ category. Such instructions have been shown to reduce lexical bias effects in lexical decision tasks (Pitt and Szostak, 2012), thus I predict a reduction in learning when attention is drawn to speech sounds. Both of these manipulations will be present in Experiments 1 and 2.

Experiments 1 and 2 differ in the atypicality of the modified /s/ category. Studies have reported greater perceptual learning when ambiguous stimuli are closer to the distribution expected by a listener than when the ambiguous stimuli are far-

ther away from expected distributions (Sumner, 2011). Words containing stimuli farther away from the target production are in general less likely to be endorsed as words, but similar effects of attention are found across word position (Pitt and Szostak, 2012). Experiment 2 contains the same manipulations to attention and lexical bias as Experiment 1, but with ambiguous stimuli farther from the target production than those used in Experiment 1. Lower rates of generalized perceptual learning are predicted for all conditions in Experiment 2.

The hypothesis of this dissertation is that the greatest perceptual learning effects should be observed when no attention is directed to the ambiguous sounds and when lexical bias is maximized. In such a case, participants should use a comprehension-oriented attentional set. If selective attention is directed to the ambiguous sounds, a more perception-oriented attentional set should be adopted with less generalization in perceptual learning as a result. Likewise, if the ambiguous sound is in a linguistically salient position with little to no lexical bias, a listener's attention should be drawn to the ambiguous sound, causing adoption of a more perception-oriented attentional set. Finally, if the ambiguous sounds are more atypical, they should be more salient to listeners regardless of lexical bias, leading again to a more perception-oriented attentional set. Regardless of the cause, adopting a perception-oriented attentional set is predicted to inhibit a generalized perceptual learning effect.

2.2 Experiment 1

In this experiment, listeners are exposed to ambiguous productions of words containing a single instance of /s/, where the /s/ has been modified to sound more /ʃ/-like. Exposure comes in the guise of a lexical decision task. In one group, the critical words have an /s/ in word-initial position (i.e., *cement*), with no /ʃ/ neighbor (a word that differs only in the sibilant; i.e., *shement*); this is referred to as the Word-initial condition. In the other group, the critical words will have an /s/ in word-medial position (*tassel*) with no /ʃ/ neighbor (*tashel*); this is referred to as the Word-medial condition. In addition, half of each group will be given instructions that the speaker has an ambiguous /s/ and to listen carefully, following Pitt and Szostak (2012).

2.2.1 Methodology

Participants

A total of 173 participants from the UBC population completed the experiment and were compensated with either $10 CAD or course credit[1]. The data from 77 nonnative speakers of English and two native speakers of English with reported speech or hearing disorders were excluded from the analyses. This left data from 94 participants for analysis. Twenty additional native English speakers participated in a pretest to determine the most ambiguous sounds. Twenty-five other native speakers of English participated for course credit in a control experiment.

Materials

Table 2.1: Filler words used in all experiments.

acorn	acrobat	antenna	apple	balloon	bamboo
buckle	butterfly	cabin	calendar	camel	campfire
candy	cockpit	collar	cowboy	cradle	cutlery
darkroom	diamond	doorbell	dryer	elephant	feather
fingerprint	garlic	goalie	gondola	graffiti	helicopter
ladder	ladle	librarian	lightning	lumber	mannequin
meadow	microwave	minivan	motel	movie	mural
napkin	omelet	painter	piano	ponytail	popcorn
referee	table	tadpole	teapot	theatre	tire
tortilla	tractor	traffic	tunnel	umbrella	weatherman

One hundred and twenty English words and 100 phonologically-legal nonwords were used as exposure materials. The set of words consisted of 40 critical items, 20 control items, and 60 filler words. Filler words and nonwords are listed in Tables 2.1 and 2.2, respectively. The control words containing /ʃ/ are given in Table 2.3. Half of the critical items had an /s/ in the onset of the first syllable

[1]The student population of the University of British Columbia has a diverse language background. In order to control for the language background of participants and to make the results of the current experiments more comparable to previous research, participants were only analyzed if they were self-reported native speakers of English. Participants were still compensated for their participation, but the data is currently unanalyzed.

Table 2.2: Filler nonwords used in Experiments 1 and 2.

apolm	arafimp	arnuff	balrop	bambany	bawapeet
bettle	bimobel	bipar	blial	brahata	danoor
darnat	deoma	follipocktel	foter	gallmit	gamtee
ganla	gippelfraw	giptern	gittle	glaple	golthin
goming	gompy	gorder	hagrant	hammertrent	hintarber
hovear	iddle	iglopad	igoldion	impomo	inoret
kempel	kimmer	kire	klogodar	kowack	lefeloo
lindel	mogmet	mopial	motpem	namittle	nartomy
nepow	neproyave	nidol	noler	nometin	nonifem
omplero	pammin	peltlon	pickpat	pidbar	pluepelai
poara	poltira	pomto	potha	prickpor	prithet
radadub	rigloriem	rinbel	rindner	ripnem	roggel
roppet	rudle	talell	talot	tankfole	tayade
teerell	tello	tepple	teygot	theely	theerheb
thorkwift	thragkole	timmer	tingora	tinogail	tirack
tirrenper	tovey	toygaw	tuckib	tuddom	tutrewy
wapteep	wekker	wogim	yovernon		

(Word-initial) and half had an /s/ in the onset of the final syllable (Word-medial). All critical tokens formed nonwords if their /s/ was replaced with /ʃ/. Half the control items had an /ʃ/ in the onset of the first syllable and half had an /ʃ/ in the onset of the final syllable. Each critical item and control item contained just the one sibilant, with no other /s z ʃ ʒ ʧ ʤ/. Filler words and nonwords did not contain any sibilants. Frequencies and number of syllables across item types are in Table 2.4

Table 2.3: Words containing /ʃ/ in all experiments.

auction	brochure	cashier	chandelier
cushion	eruption	hibernation	parachute
patient	shadow	shampoo	shareholder
shelter	shiny	shoplifter	shoulder
shovel	sugar	tissue	usher

Four monosyllabic minimal pairs were selected as test items for categorization. These minimal pairs differed only in the voiceless sibilant at the beginning of the

Table 2.4: Mean and standard deviations for frequencies (log frequency per million words in SUBTLEXus) and number of syllables of each item type

Item type	Frequency	Number of syllables
Filler words	1.81 (1.05)	2.4 (0.55)
/s/ Word-initial	1.69 (0.85)	2.4 (0.59)
/s/ Word-medial	1.75 (1.11)	2.3 (0.47)
/ʃ/ Word-initial	2.01 (1.17)	2.3 (0.48)
/ʃ/ Word-medial	1.60 (1.12)	2.4 (0.69)

word (*sack-shack*, *sigh-shy*, *sin-shin*, and *sock-shock*). Two of the pairs had a higher log frequency per million words (LFPM) from SUBTLEXus (Brysbaert and New, 2009) for the /s/ word and two had higher LFPM for the /ʃ/ word, as shown in Table 2.5.

Table 2.5: Frequencies (log frequency per million words in SUBTLEXus) of words used in categorization continua

Continuum	/s/-word frequency	/ʃ/-word frequency
sack-shack	1.11	0.75
sigh-shy	0.53	1.26
sin-shin	1.20	0.48
sock-shock	0.95	1.46

All words and nonwords were recorded by a male Vancouver English speaker in a quiet room. Critical words for the exposure phase were recorded in pairs, once normally and once with the sibilant swapped forming a nonword. The speaker was instructed to produce both forms with comparable speech rate, speech style, and prosody.

For each critical item, the word and nonword versions were morphed together in an 11-step continuum (0%-100% of the nonword /ʃ/ recording, in steps of 10%) using STRAIGHT (Kawahara et al., 2008) in Matlab. Prior to morphing, the word and nonword versions were time aligned based on acoustic landmarks, such as stop bursts, onset of F2, nasalization or frication, etc. All control items and filler words were processed and resynthesized by STRAIGHT to ensure a consistent quality

across stimulus items.

Pretest

To determine which step of each continua would be used in exposure, a phonetic categorization experiment was conducted. Participants were presented with each step of each exposure word-nonword continuum and each categorization minimal pair continuum, resulting in 495 trials (40 exposure words plus five minimal pairs by 11 steps) for each listener, blocked into exposure and categorization. Participants completed a lexical decision task for the exposure continua, responding with either "word" or "nonword" to each step of the continua. For the categorization continua, participants identified the first sound as either "s" or "sh". The experiment was implemented in E-prime (Psychology Software Tools, 2012).

Figure 2.1: Proportion of word-responses for Word-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model of the responses.

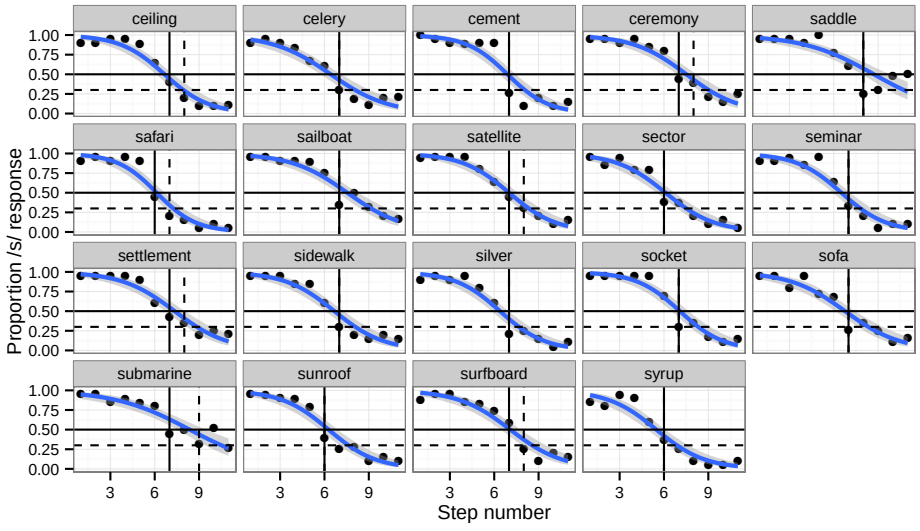


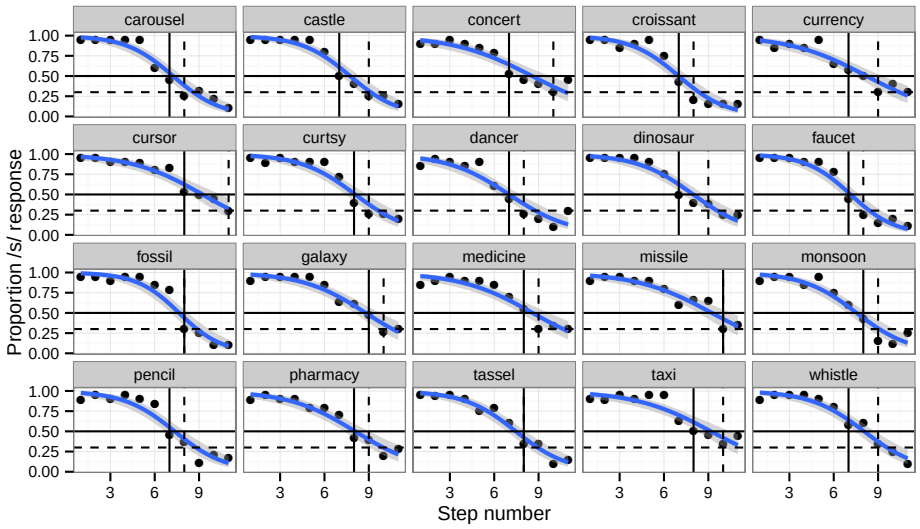
Table 2.6: Step chosen for each Word-initial stimulus in Experiment 1 and the proportion /s/ response in the pretest

Word	Step chosen	Proportion /s/ response
ceiling	7	0.40
celery	7	0.30
cement	7	0.26
ceremony	7	0.44
saddle	8	0.25
safari	6	0.45
sailboat	7	0.35
satellite	7	0.45
sector	6	0.39
seminar	7	0.33
settlement	7	0.42
sidewalk	7	0.30
silver	7	0.21
socket	7	0.30
sofa	7	0.26
submarine	7	0.45
sunroof	6	0.39
surfboard	7	0.59
syrup	6	0.37
Average	6.8	0.36

The proportion of /s/-responses (or word responses for exposure items) at each step of each continuum was calculated and the most ambiguous step chosen. The threshold for the ambiguous step for Experiment 1 was when the percentage of /s/-response dropped near 50%. The lists of steps chosen for Word-initial target stimuli are in Table 2.6 and Table 2.7, respectively. For the minimal pairs, six steps surrounding the 50% cross-over point were selected for use in the phonetic categorization task. Due to experimenter error, the continuum for *seedling* was not included in the stimuli, so the chosen step was the average chosen step for the /s/-initial words. The average step chosen for Word-initial /s/ words was 6.8 ($SD = 0.5$), and for Word-medial /s/ words the average step was 7.7 ($SD = 0.8$).

To visualize the effect of morphing on the acoustics of the sibilants and to con-

Figure 2.2: Proportion of word-responses for Word-medial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model of responses.



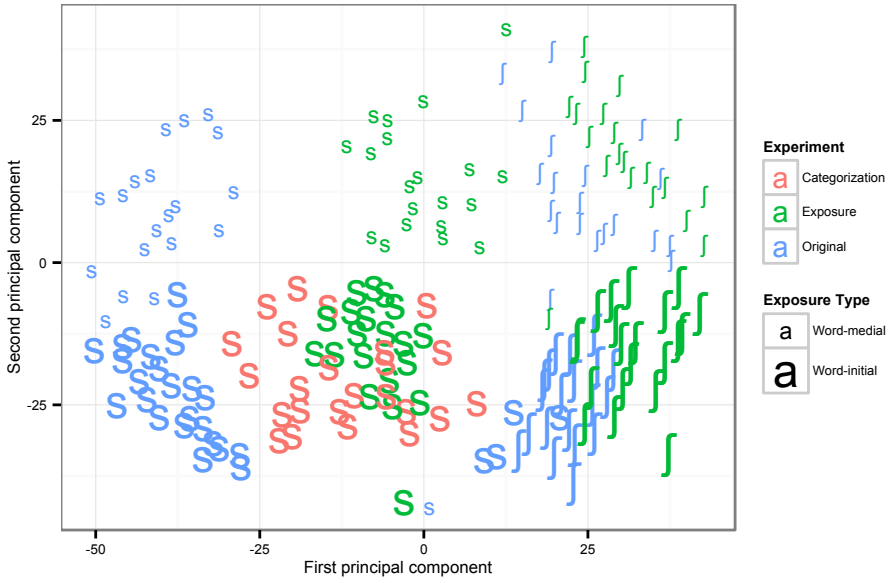
firm the desired effects, a multidimensional scaled plot of acoustic distance was constructed, similar to Mielke (2012). Using the `python-acoustic-similarity` package (McAuliffe, 2015), sibilants were transformed into arrays of mel-frequency cepstrum coefficients (MFCC), which are an auditory representation of acoustic waveforms. Pairwise distances between each sibilant production were computed via dynamic time warping to create a distance matrix of the sibilant productions. The dynamic time warping algorithm aligns time frames that are similar while allowing for time to be compressed or expanded for one of the productions. The distance returned is independent of differences in timing, but differences in order of frames are maintained. This distance matrix from the parwise calculations was then multidimensionally scaled to produce Figure 2.3. As seen there, the original, unsynthesized productions (in blue) form four quadrants based on the two princi-

Table 2.7: Step chosen for each Word-medial stimulus in Experiment 1 and the proportion /s/ response in the pretest

Word	Step chosen	Proportion /s/ response
carousel	7	0.45
castle	7	0.50
concert	7	0.53
croissant	7	0.42
currency	7	0.58
cursor	8	0.53
curtsy	8	0.40
dancer	7	0.45
dinosaur	7	0.50
faucet	7	0.45
fossil	8	0.30
galaxy	9	0.47
medicine	8	0.55
missile	10	0.30
monsoon	8	0.42
pencil	7	0.45
pharmacy	8	0.42
tassel	8	0.35
taxi	8	0.50
whistle	7	0.58
Average	7.7	0.45

pal components of the distance matrix. The first dimension is associated with the centroid frequency of the sibilant, separating /s/ tokens from /ʃ/. The second dimension separates out the word-medial sibilants (in smaller font) from the word-initial sibilants (in larger font), likely due to the different coarticulatory effects based on word position. The categorization tokens (all word-initial) predictably occupy the space between the word-initial /s/ tokens and the word-initial /ʃ/ tokens. The exposure tokens pattern as expected. Exposure /ʃ/ tokens are overlapping with the original distributions for /ʃ/ tokens. Exposure /s/ tokens are in between /s/ and /ʃ/, though word-medial /s/ tokens are closer to the original /ʃ/ distribution, reflecting the difference in average stimuli step chosen in Tables 2.6 and 2.7.

Figure 2.3: Multidimensional scaling of the acoustic distances between the sibilants of original productions, categorization tokens and the exposure tokens in Experiment 1. Categorization and exposure tokens were synthesized from the original productions using STRAIGHT (Kawahara et al., 2008).



Experiment design

Participants were assigned to one of four groups from a 2x2 between-subject factorial design. The first factor was the position of the ambiguous sibilant in the exposure words (Exposure Type: Word-initial versus Word-medial) and the second factor was whether participants were given additional instructions about the sibilant (Attention: Attention versus No Attention). Two of the groups of participants were exposed to only critical items that began with /s/ (Word-initial) and the other two were exposed to only critical items that had an /s/ in the onset of the final syllable (Word-medial). This gave a consistent 200 trials in all exposure phases with identical control and filler items for all participants. Participants in half the groups (Attention) received additional instructions that the speaker's "s"

sounds were sometimes ambiguous, and to listen carefully to ensure correct responses in the lexical decision. Participants in the control experiment completed only the categorization task.

Procedure

Participants in the experimental conditions completed an exposure task and a categorization task in E-Prime (Psychology Software Tools, 2012). Exposure was a lexical decision task, where participants heard auditory stimuli and were instructed to respond with either "word" if they thought what they heard was a word or "nonword" if they did not think it was a word. The buttons corresponding to "word" and "nonword" were counterbalanced across participants. Trial order was pseudorandom. Stimuli containing sibilants (/s/ or /ʃ/) did not appear in the first six trials or next to each other (following Reinisch et al., 2013). For each trial, a blank screen was shown for 500 ms, and then the two responses and their corresponding buttons on the button box were shown (i.e. "word" and "1" on one side of the screen and "nonword" and "5" on the other side of the screen). The auditory stimulus was played 500 ms following the presentation of the response options. Participants had 3000 ms from the onset of the auditory stimulus to respond. Participants were given feedback whether a response was detected in the 3000 ms window. This feedback did not include accuracy or response time information and was shown for 500 ms before the following trial began. Every 50 trials participants were given a break and the next trial did not start until the participant pressed a button.

In the categorization task, participants heard an auditory stimulus and had to categorize it as one of two words, differing only in the onset sibilant, i.e. *sin* or *shin*. The buttons corresponding to the words were counterbalanced across participants. The six most ambiguous steps of the minimal pair continua were used with seven repetitions each, giving a total of 168 trials. Each trial proceeded similarly to exposure. A blank screen was displayed for 500 ms, followed by the response screen for 500 ms (i.e. "sin" and "1" on one side, "shin" and "5" on the other) before the auditory stimulus was presented. Participants had 3000 ms from the onset of the auditory stimulus to respond and feedback about whether the response was detected was shown for 1500 ms. Participants were given a break every 40

trials, except after 160 trials, as that would leave eight trials in the rest of the experiment.

To remove experimenter interaction between exposure and categorization, participants were given oral instructions explaining both tasks at the beginning of the experiment. Written instructions were presented to participants at the beginning of each task as well. The instructions for the exposure task given to participants assigned to an Attention condition included explicit reference to the modified sibilants. Participants were told that "this speaker's 's' sound is sometimes ambiguous" and instructed to "listen carefully so as to choose the correct response."

Analysis

Perceptual learning effects are assessed through logistic mixed-effects models of the categorization task data. Responses were coded as 1 for /s/ responses and 0 for /ʃ/ responses. Positive significant estimates therefore indicate higher likelihood of /s/ response across categorization. Thus, positive significant effects are indicative of perceptual learning, as higher likelihood of /s/ response is associated with an expanded /s/ category.

Deviance contrast coding is used for all two-level independent variables, so the intercept of the model represents the grand mean. Main effects for factors are calculated with other factors held at their average value, rather than at an arbitrary reference level. For any factors that have three levels, treatment (dummy) contrasts are used with an appropriate reference level to aid interpretation. Numeric independent variables were centered prior to inclusion in models. Although continua steps are discrete (i.e., Step 1 and Step 2, but no intermediate tokens), it is entered as a numeric variable in the models to reduce the complexity of models. Graphs of categorization results show continua step as categorical factor to aid interpretation.

For categorization models, Continuum was a random effect. However, there were only four minimal pair continua used in the categorization, so the random effect status may not be warranted. The estimates for the continua effects are likely not very reliable, but differences between continua are not the principle question being investigated. Use of a by-Continuum random effect structure with maximal random slopes allowed for estimation of the fixed effects that are not driven by one

particular minimal pair continuum.

2.2.2 Results

Control experiment

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses (following Reinisch et al., 2013). A logistic mixed effects model was fit with Subject and Continuum as random effects and Step as a fixed effect with by-Subject and by-Continuum random slopes for Step. The intercept was not significant ($\beta = 0.43, SE = 0.29, z = 1.5, p = 0.13$), indicating that control participants did not differ significantly from the pretest participants. Step was significant ($\beta = -2.61, SE = 0.28, z = -9.1, p < 0.01$), with higher steps (more /ʃ/-like) responded to more as /ʃ/. Results from the control experiment are shown in Figure 2.6 and all other categorization results as a reference point for interpreting the figures.

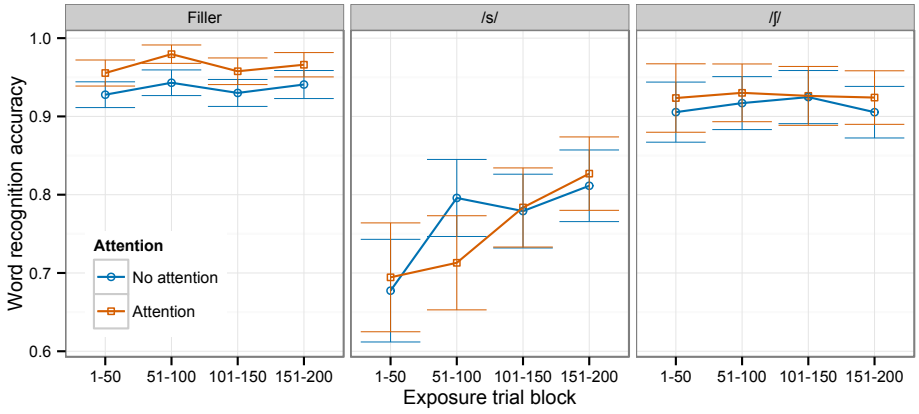
Exposure

Performance on the exposure task was high overall: 92% of the filler words were correctly accepted and 89% of nonwords were correctly rejected. Trials with nonword stimuli and responses faster than 200 ms or slower than 2500 ms were excluded from further analysis. A logistic mixed-effects model with accuracy as the dependent variable was fit with fixed effects for Trial (0-200), Trial Type (Filler, /s/, and /ʃ/), Attention (No Attention and Attention), Exposure Type (Word-Initial and Word-Medial), and their interactions. Trial Type was coded using treatment (dummy) coding, with Filler as the reference level. Deviance contrast coding was used for Exposure Type (Word-initial = 0.5, Word-medial = -0.5) and Attention (No attention = 0.5, Attention = -0.5). The random effect structure was as maximally specified as possible with random intercepts for Subject and Word. By-Subject random slopes were specified for Trial, Trial Type, and their interactions. By-Word random slopes were specified for Attention, Exposure Type, and their interactions

A significant fixed effect was found for Trial Type of /s/ versus Filler ($\beta = -2.13, SE = 0.31, z = -6.8, p < 0.01$), as participants were less likely to endorse

words containing the modified /s/ category as compared to filler words. However, there was a significant interaction between Trial and Trial Type of /s/ versus Filler ($\beta = -0.45, SE = 0.14, z = 3.1, p = 0.01$), so the differences in accuracy between words with /s/ and filler words diminished over time. Participants adapted to the speaker's /s/ over the course of the experiment. There was also a significant main effect of Attention ($\beta = -0.57, SE = 0.28, z = -2.0, p = 0.04$), indicating that participants were more accurate at identifying words in the Attention condition compared to the No Attention condition. However, there was a marginal interaction between Attention and Trial Type of /s/ versus Filler ($\beta = 0.72, SE = 0.39, z = 1.8, p = 0.06$), suggesting that attention only increased accuracy for words not containing the modified /s/ category. Figure 2.4 shows within-subject mean accuracy across exposure, with Trial in four blocks.

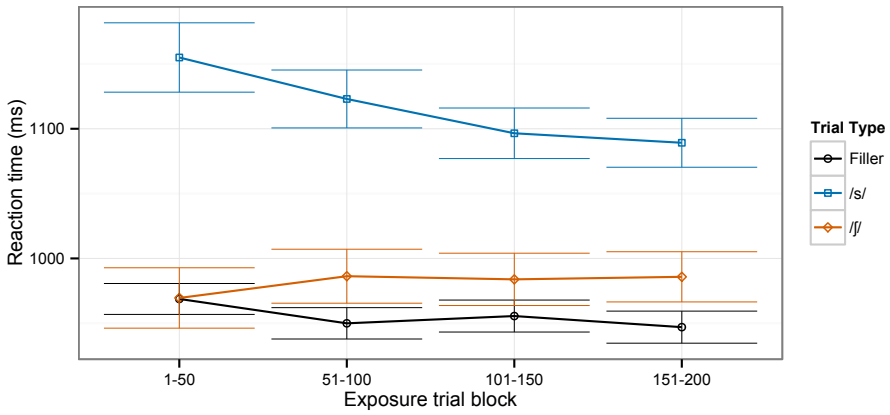
Figure 2.4: Within-subject mean accuracy for words in the exposure phase of Experiment 1, separated out by Trial Type (Filler, /s/, and /ʃ/). Error bars represent 95% confidence intervals.



A linear mixed-effects with logarithmically-transformed reaction time as the dependent variable was fit with identical fixed effect and random effect structure as the logistic model for accuracy. Significant effects were found for Trial Type of /s/ versus Filler ($\beta = 0.71, SE = 0.07, t = 10.8$) and the interaction between Trial and Trial Type of /s/ versus Filler ($\beta = -0.08, SE = 0.02, t = -3.1$). These

effects follow the pattern found in the accuracy model, where participants begin with slower reaction times to words with /s/, but over time this difference between words /s/ and filler words lessens. Figure 2.5 shows within-subject mean reaction time across exposure, with Trial in four blocks.

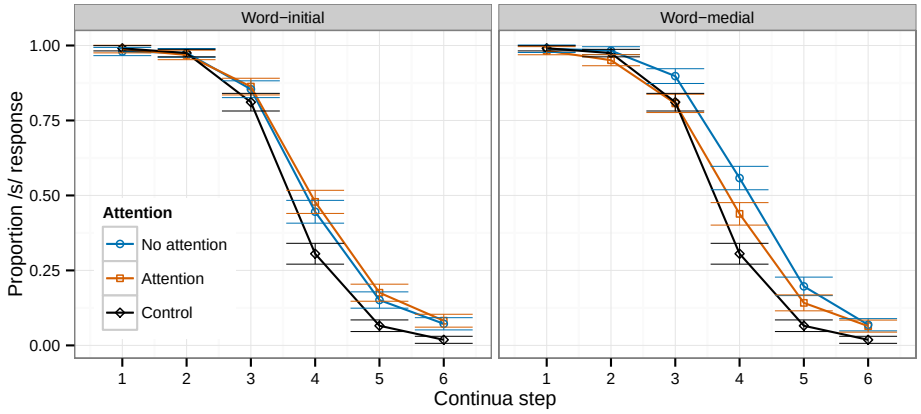
Figure 2.5: Within-subject mean reaction time to words in the exposure phase of Experiment 1, separated out by Trial Type (Filler, /s/, and /ʃ/). Error bars represent 95% confidence intervals.



Categorization

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. Two participants were excluded because their initial estimated cross-over point for the continuum lay outside of the 6 steps presented. A logistic mixed effects model was constructed with Subject and Continuum as random effects and a by-Subject random slope for Step and by-Continuum random slopes for Step, Attention, Exposure Type, and their interactions. Fixed effects for the model were Step, Exposure Type, Attention, and their interactions. Deviance contrast coding was used for Exposure Type (Word-initial = 0.5, Word-medial = -0.5) and Attention (No attention = 0.5, Attention = -0.5). An /s/ response was coded as 1 and an /ʃ/ response as 0.

Figure 2.6: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 1. Error bars represent 95% confidence intervals.



There was a significant effect for the intercept ($\beta = 0.76, SE = 0.22, z = 3.3, p <$ 0.01), indicating that participants categorized more of the continua as /s/ in general. This is evidence of learning compared to participants in the control experiment. There was also a significant main effect of Step ($\beta = -2.14, SE = 0.15, z = -14.2, p < 0.01$), and a significant interaction between Exposure Type and Attention ($\beta = -0.93, SE = 0.45, z = -2.04, p = 0.04$). The results are visualized in Figure 2.6. When exposed to a modified /s/ category at the beginning of words, participants show a general expansion of the /s/ category with no difference in behavior induced by the attention manipulation. However, when the exposure is to ambiguous /s/ tokens later in the words, we can see differences in behavior beyond the general /s/ category expansion. Participants who were not warned of the speaker's ambiguous tokens categorized more of the continua as /s/ compared to those who were warned of the speaker's ambiguous /s/ productions.

2.2.3 Discussion

The condition that showed the largest perceptual learning effect was the one most biased toward a comprehension-oriented attentional set. Participants exposed to the modified /s/ category in the middle of words and with no explicit instructions about /s/ had larger perceptual learning effects than any of the other conditions. The other conditions showed roughly equivalent sizes of perceptual learning, suggesting that there was not a compounding effect of explicit attention and word position. That is, the comprehension-oriented nature of the primary task still exerts an effect on attentional set selection, and a significant perceptual learning effect was found on novel words.

The findings of this experiment do not support the predictions of a purely gain-based mechanism for attention, such as the one posited by Clark (2013). If attention functioned as a gain mechanism – that is, increasing the weight of error signals generated by mismatches between expectations and incoming signals – we should expect to see greater perceptual learning when listeners were instructed to attend to the speaker's /s/ sounds. Instead, the opposite was found. Participants told to attend to the speaker's /s/ sounds showed smaller perceptual learning effects. The nature and sentiment of the instructions may affect the outcome of attention. In this experiment, the instructions regarding /s/ were phrased to suggest that the ambiguity of the speaker's "s" could harm accuracy, a negative sentiment. If the instructions about the speaker's "s" were more positive, such as giving an explanation for the cause of ambiguity, then a different pattern might be observed. For a gain-based mechanism, however, positive or negative sentiment in attention is not predicted to affect attention, but rather attention always increasing the gain of error signals. If sentiment of the instructions does change behavior, then it would be another mark against a gain mechanism.

In addition to the perceptual learning effects of the categorization phase, the exposure phase also demonstrates learning. In the initial trials, words with a modified /s/ are responded to more slowly and less accurately, but over the course of exposure, both reaction times and accuracy approach those of filler and unmodified /ʃ/ words. Interestingly, only the attention manipulation had an effect on exposure performance, with participants attending to the /s/ category responding more accu-

rately overall. Exposure Type did not significantly influence accuracy or reaction time in the exposure task.

Much of the literature on perceptual learning in speech perception focuses on the issue of generalization and specificity. For instance, listeners have been shown to generalize across speakers more if the exposure speaker's modified category happens to be within the range of variation of the categorization speaker's stimuli (Eisner and McQueen, 2005; Kraljic and Samuel, 2005). Additionally, many perceptual learning studies artificially enhance the similarity between exposure tokens and categorization tokens, such as splicing the maximally ambiguous step of the categorization continuum into exposure words (Norris et al., 2003). Because exposure-specificity plays such a large role in perceptual learning, it is natural to consider whether the greater perceptual learning effects in some conditions arise due to greater similarity to the exposure stimuli. However, as shown in Figure 2.3, Word-medial exposure tokens are acoustically farther from the categorization tokens than the Word-initial exposure tokens. Even if auditory similarity of /s/ across exposure and categorization played any role, it was still overridden by the experimental manipulations.

In this experiment I used a similar method for exposure stimuli selection as Reinisch et al. (2013), but used a threshold of 50% word response rate in the pretest as the cutoff rather than 70%. With their 70% stimuli, Reinisch and colleagues report word endorsement rates that consistently exceeded 85%. In contrast, Experiment 1 used 50% as the threshold and had correspondingly lower word endorsement rates (mean = 76%, SD = 22%). Despite the lower word endorsement rates and the less canonical stimuli used, perceptual learning effects remained robust. This raises the question: can perceptual learning occur from a modified category even more atypical than the one used in this experiment? More atypical categories should be more salient and induce a more perception-oriented attentional set, and therefore result in smaller perceptual learning effects. In Experiment 2, we test whether a comprehension-oriented attentional set can be maintained despite the category atypicality triggering perception-oriented attentional sets.

2.3 Experiment 2

Experiment 2 uses stimuli that are farther from the canonical productions of the critical exposure tokens containing /s/.

2.3.1 Methodology

Participants

A total of 127 participants from the UBC population completed the experiment and were compensated with either $10 CAD or course credit. The data from 31 nonnative speakers of English were excluded from the analyses. This left data from 96 participants for analysis.

Materials

Experiment 2 used the same items as Experiment 1, except that the step along the /s/-/ʃ/ continua chosen as the ambiguous sound had a different threshold. For this experiment, 30% identification as the /s/ word was used the threshold. The average step chosen for /s/-initial words was 7.3 ($SD = 0.8$), and for /s/-medial words the average step was 8.9 ($SD = 0.9$). The list of steps chosen for Word-initial and Word-medial target stimuli are in Tables 2.8 and 2.9, respectively. Note that for several stimuli, the same steps are used for both Experiment 1 and 2. There were large jumps in proportion /s/ response between steps for the continua for those stimuli. However, the key aspect of the stimuli is the distribution of the /s/ category as a whole, and not the individual steps.

Multidimensional scaling was employed to assess the distributions of the stimuli used in Experiment 2. A similar pattern is found for Experiment 2 as Experiment 1. The axes remain the same as before, with the first dimension corresponding to differences between sibilants, and the second dimension corresponding to differences in word position. The original productions, categorization tokens, and /ʃ/ tokens in Figure 2.7 are identical to those shown in Figure 2.3, but the exposure token distribution is shifted towards the /ʃ/ distribution. In the Word-medial position, the distributions of /s/ and /ʃ/ are close to overlapping, and in the Word-initial position, they are still separated, but closer than in the stimuli for Experiment 1.

Table 2.8: Step chosen for each Word-initial stimulus in Experiment 2 and the proportion /s/ response in the pretest

Word	Step chosen	Proportion /s/ response
ceiling	8	0.20
celery	7	0.30
cement	7	0.26
ceremony	8	0.39
saddle	8	0.25
safari	7	0.21
sailboat	7	0.35
satellite	8	0.30
sector	6	0.39
seminar	7	0.33
settlement	8	0.35
sidewalk	7	0.30
silver	7	0.21
socket	7	0.30
sofa	7	0.26
submarine	9	0.32
sunroof	6	0.39
surfboard	8	0.25
syrup	6	0.37
Average	7.3	0.30

Procedure

The procedure and instructions were identical to those of Experiment 1.

Analysis

Response data and factors were transformed and analyzed in the same way as in Experiment 1.

Table 2.9: Step chosen for each Word-medial stimulus in Experiment 2 and the proportion /s/ response in the pretest

Word	Step chosen	Proportion /s/ response
carousel	8	0.25
castle	9	0.25
concert	10	0.30
croissant	8	0.20
currency	9	0.30
cursor	11	0.30
curtsy	9	0.26
dancer	8	0.26
dinosaur	9	0.39
faucet	8	0.25
fossil	8	0.30
galaxy	10	0.26
medicine	9	0.30
missile	10	0.30
monsoon	9	0.15
pencil	8	0.37
pharmacy	9	0.39
tassel	8	0.35
taxi	10	0.35
whistle	9	0.35
Average	8.9	0.29

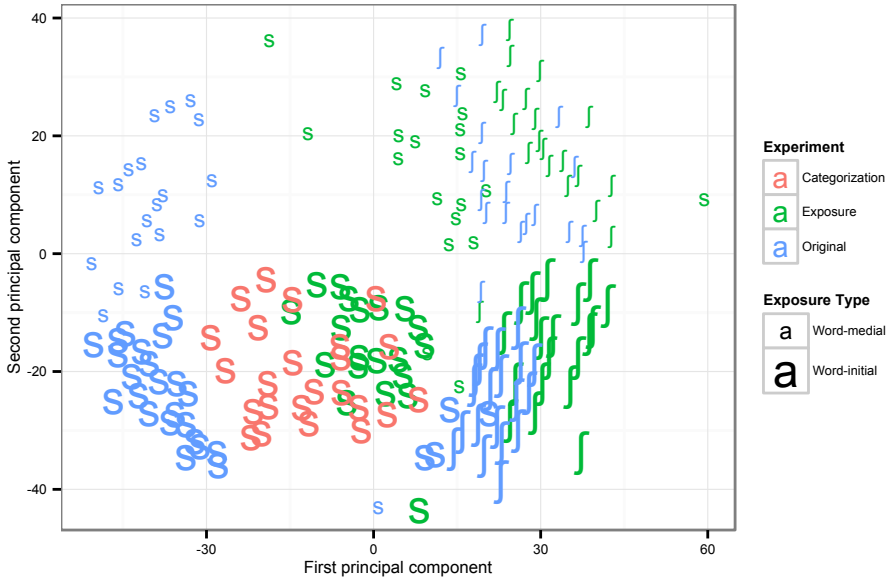
2.3.2 Results

Exposure

Trials with nonword stimuli and responses faster than 200 ms or slower than 2500 ms were excluded from analysis. Performance on the exposure task was as high as in Experiment 1, with accuracy on filler trials averaging 92%. A logistic mixed-effects model with accuracy as the dependent variable and a linear mixed-effects model with reaction time (logarithmically-transformed) as the dependent variable were fit with identical specifications as Experiment 1.

In the logistic mixed-effects model of accuracy, a significant fixed effect was

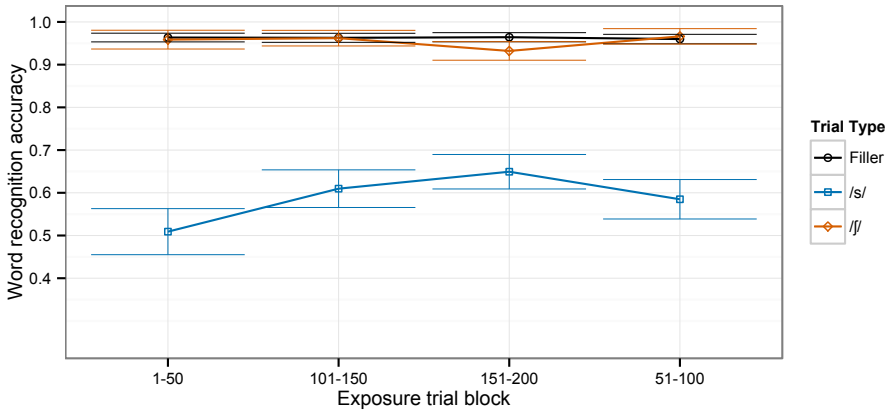
Figure 2.7: Multidimensional scaling of the acoustic distances between the sibilants of original productions, categorization tokens and the exposure tokens in Experiment 2. Categorization and exposure tokens were synthesized from the original productions using STRAIGHT (Kawahara et al., 2008).



found for Trial Type of /s/ versus Filler ($\beta = -3.56, SE = 0.31, z = -11.4, p < 0.01$), with participants less likely to respond that an item was a word if it contained the modified /s/ category. There was a significant interaction between Trial and Trial Type of /s/ versus Filler ($\beta = -0.29, SE = 0.11, z = 2.5, p < 0.01$) and between Trial and Trial Type of /ʃ/ versus Filler ($\beta = -0.33, SE = 0.16, z = -2.0, p = 0.04$). These interactions indicate that participants became more likely to endorse words with modified /s/ productions over time, but also became less accurate on words containing /ʃ/. Figure 2.8 shows within-subject mean accuracy across exposure, with Trial in four blocks.

In the linear mixed-effects model of reaction time, significant effects were found for Trial Type of /s/ versus Filler ($\beta = 0.94, SE = 0.07, t = 14.4$), indicat-

Figure 2.8: Within-subject mean accuracy in the exposure phase of Experiment 2, separated out by Trial Type (Filler, /s/, and /ʃ/). Error bars represent 95% confidence intervals.

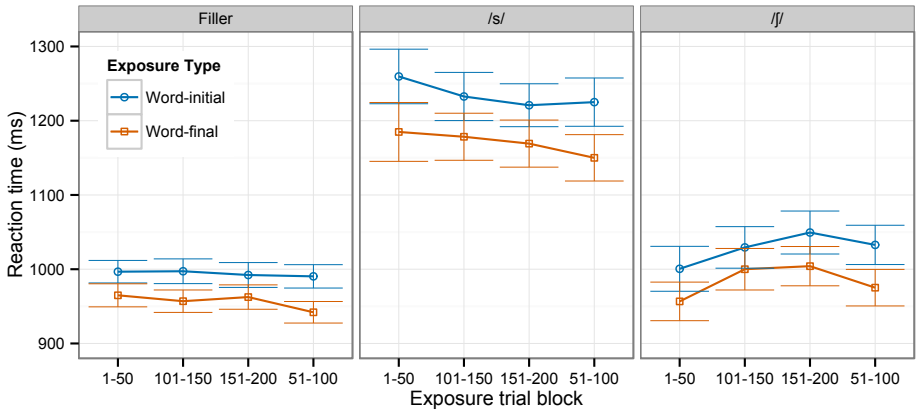


ing that reaction times were slower for words containing the modified /s/ category. Also significant was the interaction between Trial and Trial Type of /ʃ/ versus Filler ($\beta = 0.07, SE = 0.02, t = 3.4$). However, there was no interaction between Trial and Trial Type of /s/ versus Filler ($\beta = -0.02, SE = 0.02, t = -0.8$). This indicates that reaction time remained relatively stable for words containing the modified /s/ category, but lengthened for words containing the /ʃ/ control. There was a marginal effect for Trial Type of /ʃ/ versus Filler ($\beta = 0.14, SE = 0.07, t = 1.9$), indicating that words with /ʃ/ tended to be responded to more slowly than filler times. Finally, there was a marginal effect of Exposure Type ($\beta = 0.17, SE = 0.09, t = 1.9$), indicating that words in the Word-medial condition tended to be responded to faster. Figure 2.9 shows within-subject mean reaction time across exposure, with Trial in four blocks.

Categorization

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. Two participants were excluded because their initial esti-

Figure 2.9: Within-subject mean reaction time in the exposure phase of Experiment 2, separated out by Trial Type (Filler, /s/, and /ʃ/). Error bars represent 95% confidence intervals.



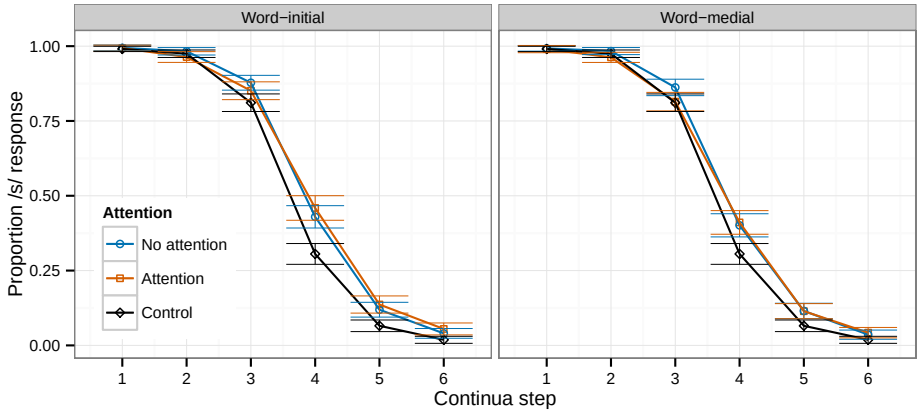
mated cross-over point for the continuum lay outside of the 6 steps presented. A logistic mixed effects model was constructed with identical specification as Experiment 1.

There was a significant effect for the Intercept ($\beta = 0.60, SE = 0.26, z = 2.3, p = 0.02$), indicating that participants categorized more of the continua as /s/ in general. There was also a significant main effect of Step ($\beta = -2.51, SE = 0.19, z = -13.1, p < 0.01$). Unlike in Experiment 1, there were no other significant effects, suggesting that participants across conditions had similar perceptual learning effects. These results are shown in Figure 2.10

2.4 Grouped results across experiments

The data from Experiment 1 and Experiment 2 were pooled and analyzed identically as above, but with Experiment and its interactions as additional fixed effects to directly assess the effect of category atypicality. In the logistic mixed effects model, there were significant main effects for Intercept ($\beta = 1.00, SE = 0.36, z = 2.7, p < 0.01$) and Step ($\beta = -2.64, SE = 0.21, z = -12.1, p < 0.01$),

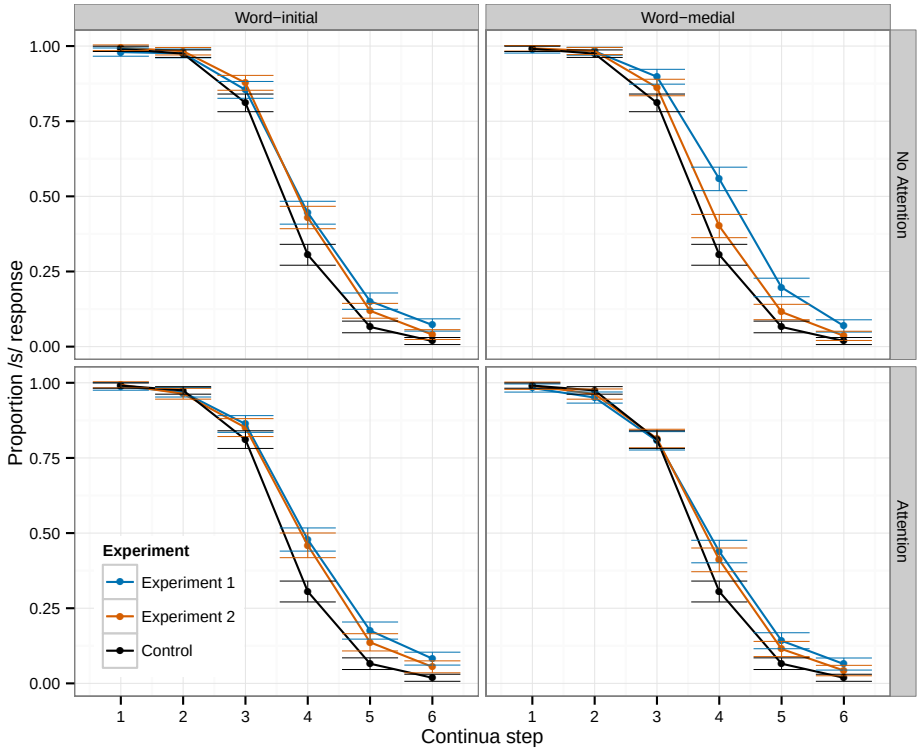
Figure 2.10: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 2. Error bars represent 95% confidence intervals.



a significant two-way interaction between Experiment and Step ($\beta = 0.51, SE = 0.20, z = 2.5, p = 0.01$), and a marginal four-way interaction between Step, Exposure Type, Attention and Experiment ($\beta = 0.73, SE = 0.42, z = 1.7, p = 0.08$). These results can be seen in Figure 2.11. The four-way interaction can be seen in Word-medial/No Attention conditions across the two experiments, where Experiment 1 has a difference between the Attention and No Attention condition, but Experiment 2 does not. The two-way interaction between Experiment and Step and the lack of a main effect for Experiment suggests that while the category boundary was not significantly different across experiments, the slope of the categorization function was.

In previous research, a link has been shown between the proportion of word endorsement for exposure tokens and the size of perceptual learning effects (Scharenborg and Janse, 2013). Listeners who endorsed more of exposure tokens as words showed a larger perceptual learning effect. To assess such a link in the current experiments, a logistic mixed-effects model was constructed identically as above. However, participants' word endorsement rate of target /s/ words were included as an additional fixed effect, along with its interactions with all other fixed effects.

Figure 2.11: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 1 and Experiment 2. Error bars represent 95% confidence intervals.

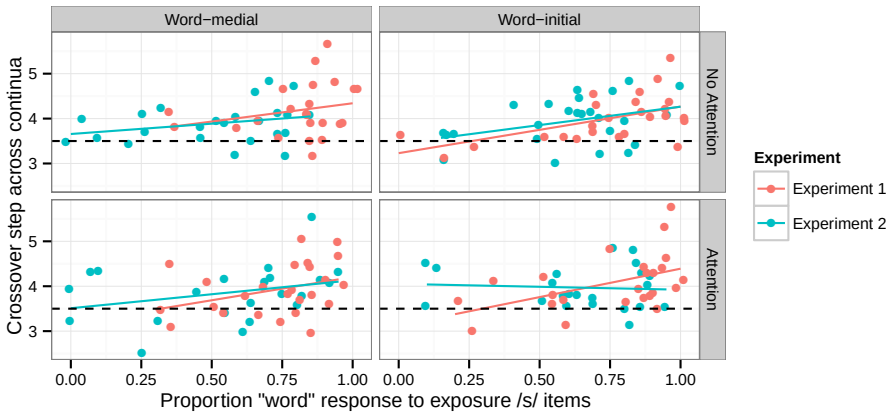


Word endorsement rate was calculated as the ratio of the number of word responses by the total number of /s/ trials. Prior to inclusion in the model, an arcsine transformation was performed on the word endorsement rates. Word endorsement rate was significant ($\beta = 1.55, SE = 0.38, z = 4.1, p < 0.01$), finding the same effect as previous work. Participants who endorsed more exposure tokens showed larger perceptual learning effects. Word endorsement rate significantly interacted with Step ($\beta = 0.63, SE = 0.24, z = 2.6, p < 0.01$) and was involved in a marginal interaction with Step, Attention and Experiment ($\beta = -1.79, SE = 0.94, z = -1.8, p = 0.06$).

To better investigate the nature of the four-way interaction, word endorsements

were correlated with estimated cross-over points by participant. Cross-over points occur when a participant's perception switches from predominantly /s/ to predominantly /ʃ/. The cross-over point was determined from the Subject random intercept and the by-Subject random slope of Step in a simple model containing only those random effects, similar by-Continuum random effects, and a fixed effect for Step (Kleber et al., 2012). Scatter plots of word endorsement rate and cross-over point across all experimental conditions are shown in Figure 2.12. In general there is a positive correlation between word endorsement rate in the exposure phase and the cross-over point from /s/ to /ʃ/ in the categorization phase. Participants in Experiment 1 who were exposed to more typical /s/ stimuli showed a stronger correlation across conditions than participants in Experiment 2, who were exposed to a more atypical /s/ category. An analysis of word endorsement rates across Exposure Type, Attention, and Experiment revealed only a significant difference in endorsement rates for Experiment ($F(1,187) = 26.8, p < 0.01$). Experiment 1 had a mean word endorsement rate of 75% (SD = 23%) and Experiment 2 had a mean endorsement rate of 58% (SD = 27%).

Figure 2.12: Correlation of cross-over point in categorization with the proportion of word responses to critical items containing an ambiguous /s/ token in Experiments 1 and 2.



In Experiment 1, the strongest correlations between word endorsement rate and

cross-over point are seen in the Word-initial conditions (Attention: $r = 0.46, t(22) = 2.4, p = 0.02$; No Attention: $r = 0.45, t(22) = 2.4, p = 0.02$), with the next strongest, and more marginal, correlation in Word-medial/Attention condition ($r = 0.39, t(23) = 2.0, p = 0.06$). The condition for which the most perceptual learning was observed (Word-medial/No Attention) actually has the weakest relationship ($r = 0.32, t(21) = 1.5, p = 0.13$).

In Experiment 2, the strongest correlation is in the Word-initial/No Attention condition ($r = 0.40, t(23) = 2.1, p = 0.05$), with two trending correlations for the Word-medial conditions (Attention: $r = 0.33, t(20) = 1.6, p = 0.12$; No Attention: $r = 0.27, t(22) = 1.3, p = 0.20$). Finally, the correlation for the Word-initial/Attention condition is not significant ($r = -0.05, t(20) = -0.2, p = 0.82$).

2.5 General discussion

The perceptual learning effects found in Experiment 1 and 2 align with either comprehension-oriented or perception-oriented attentional sets. The perception-oriented attentional sets are predicted to exhibit less generalization, similar to what is seen in the psychophysics literature and in visually-guided perceptual learning in speech perception (Reinisch et al., 2014). In support of this, participants in perception-oriented conditions of Experiment 1 (i.e., Attention conditions and Word-initial conditions) showed uniform and modest amounts of perceptual learning. Those in Experiment 2, who were triggered to use a perception-oriented set based on the category atypicality of the stimuli, showed similar modest levels of perceptual learning. Participants that were not exposed to any triggers towards perception-oriented attentional sets (Experiment 1/ No Attention/ Word-medial) were predicted to use a more comprehension-oriented attentional set which aligns with the task performed. These participants showed a substantially larger perceptual effect than those biased towards perception-oriented attentional sets.

Compared to Experiment 1, Experiment 2 had a weaker correlation between critical word endorsement rates and cross-over boundary points. This suggests that although the stimuli used in Experiment 2 were farther from the canonical production, they did not shift the category boundary as much as the stimuli in Experiment 1. While neither attention nor position of the ambiguous sound had an effect on

the correlation, the distance from the canonical production did. This potentially suggests that the degree to which a category is shifted is inversely related to the token's distance to the mean.

One condition in Experiment 1 did not have a strong correlation between word response rate and cross-over point. This condition (No Attention/Word-medial exposure) had the largest perceptual learning effect, as well. The lack of correlation in precisely this condition falls out from the proposed attention mechanism. Comprehension-oriented attentional sets are proposed to update higher, more abstract linguistic representations. Initial endorsements might shift the boundary more than later endorsements under such an attentional set, which would result in a non-linear relationship between endorsement rate and cross-over point. Lexically-guided perceptual learning is typically induced with relatively few tokens, usually 20 of 200 total tokens (Norris et al., 2003; Reinisch et al., 2013), but as few as 10 modified tokens have been shown to cause perceptual learning (Kraljic et al., 2008b). Under perception-oriented attentional sets, the relationship between endorsement rate and cross-over point may be more linear, with equal updating per endorsement, but each individual instance contributes less than initial comprehension-oriented endorsements. Visually-guided perceptual learning generally uses hundreds of target tokens with no fillers (Vroomen et al., 2007; Reinisch et al., 2014).

The correlation between word response rate in the exposure phase and the category boundary in the categorization phase across both experiments has two possible explanations. In a causal interpretation between exposure and categorization, as each ambiguous sound is processed and errors propagate, the distribution for that category (for that particular speaker) is updated. Participants who processed more of the ambiguous sound as an /s/ updated their perceptual category for /s/ more. This explanation fits within a Bayesian model of the brain (Clark, 2013) or a neo-generative model of spoken language processing (Pierrehumbert, 2002). A non-causal story is also plausible: the correlation may reveal individual differences on the part of the participants, where some participants are more adaptable or tolerant of variability than others. These more tolerant listeners then show greater degrees of perceptual adaptation. Individual differences in attention-switching control have previously been found to affect perceptual learning (Scharenborg et al.,

2014), which supports a non-causal interpretation as well.

As mentioned in the discussion for Experiment 1, the findings do not support a simple gain mechanism for attention (contra Clark, 2013). In Yeshurun and Carrasco (1998), attention to areas with finer spatial resolution caused observers to miss larger patterns. If attention simply boosted the error signal, attention of all kinds should always be beneficial to perceptual learning. The findings of these two experiments supports a larger role for attention in a predictive coding framework, which been previously noted in Block and Siegel (2013). The propagation-limiting attention mechanism proposed in this dissertation explains both the findings in the visual domain and the current findings. Attention to a level of representation causes errors between expectations and observed signals to be resolved and updated at that level. If attention is more oriented towards comprehension, errors can be propagated to a higher, more abstract level of linguistic representation before updating expectations.

A lexical decision task by default biases a participant towards a comprehension-oriented attentional set. The experimental manipulations promoted perception-oriented attentional sets that attenuated generalized perceptual learning effects. To fully examine the use of perception- and comprehension-oriented attentional sets in perceptual learning, manipulations that induce comprehension-oriented attentional sets are necessary. In the following chapter, such a manipulation is implemented through increasing linguistic expectations with semantic predictability.

Chapter 3

Cross-modal word identification

3.1 Motivation

The largest perceptual learning effect in this dissertation was found in Experiment 1 in the No Attention/ Word-medial condition, which is the condition that was the least likely to promote a perception-oriented attentional set in listeners. The lexical decision task is a comprehension-oriented task, so the comprehension-oriented attentional set is the default attentional set. Participants with no manipulation promoting a perception-oriented attentional set would have maintained this default attentional set. The experiment in this chapter examines perceptual learning in larger sentence contexts, as opposed to lexically guided perceptual learning in single word paradigms. Using sentences ending with final target words that contain a modified /s/ category, semantic predictability is used in conjunction with lexical bias to boost linguistic expectations. The linguistic expectation exploited in Chapter 2 was a lexical expectation. Hearing part of a word increases a listener's expectation for hearing the rest of that word. But as the words are presented in isolation in the lexical decision task, all words have equal likelihood of occurring. No particular expectations are present prior to hearing the initial sounds of the word. In a sentence context, however, expectations of a particular word can be boosted by the words preceding it. If the expectation for a word is increased, the expectations for the sounds within it would be increased as well.

For our purposes, semantic predictability refers to how predictable the final

word in a sentence is (Kalikow et al., 1977). Example (1) is a high predictability sentence and Example (2) is an unpredictable sentence. Although the final word is the same in both, the preceding sentence cues the final word in the predictable sentence, but not in the unpredictable one. Semantic predictability does not reference formal models of semantics explicitly, but is rather more about world knowledge.

(1) The cow gave birth to the calf.

(2) She is glad Jane called about the calf.

In general, high predictability sentences contain less signal information, but are easier to process and understand. For example, semantic predictability in production studies is associated with phonetic reduction (particularly duration of words and sounds) independent of lexical factors like frequency and neighborhood density (Scarborough, 2010; Clopper and Pierrehumbert, 2008). In speech perception work, semantic predictability and lexical bias have been found to have similar effects on phoneme categorization (Connine, 1987; Borsky et al., 1998). Sentences with higher semantic predictability are more intelligible in noise, particularly for native speakers (Kalikow et al., 1977; Mayo et al., 1997; Fallon et al., 2002; Bradlow and Alexander, 2007, and others). Similarly, in phoneme restoration tasks, semantic predictability increases the bias for a listener to hear a complete word, which may account for the increased intelligibility. However, this increased bias is coupled with an increased sensitivity in detecting missing sounds for semantically predictable words (Samuel, 1981).

Samuel (1981) proposes that high predictability sentences place a lower cognitive load on the perceptual system. The lower cognitive load allows for more cognitive resources to be allocated to the primary perception-oriented task, resulting in greater perceptual sensitivity. Mattys and Wiget (2011) manipulated cognitive load through easier or harder concurrent visual search tasks during a phoneme categorization task. Mirroring the phoneme restoration results, Mattys and colleagues found greater perceptual sensitivity in conditions with lower cognitive load. In both of these cases, the goal of the listener was oriented towards perception. A lower cognitive load on the perceptual system may allow more cognitive resources (including attention) to be allocated to perception. In a comprehension-oriented

task, lower cognitive load would not necessarily always result in greater perceptual sensitivity. If a listener's end goal is not perception of a specific production of a speech sound, then performance on the task would not necessarily be increased by attending further to perception.

There are many possible outcomes for perceptual learning in this experiment. Many theoretical frameworks do not make explicit predictions about how the perceptual system will be updated in the context of full sentences. Most models of speech perception end at perception of words. In a sentence like Example (1), perceiving individual words is likely not the goal. For instance, independently perceiving the word *to* does not aid in comprehending the meaning of the sentence. Instead, comprehension is likely more oriented towards the relations between concepts and perceiving phrases or multiword chunks. If the perceived chunk is larger than a word, are the fine details still as faithfully encoded as they are for words in isolation? Even if the fine details are encoded, are they reliable enough evidence for perceptual learning? The experiment reported here takes a first pass at answering some of these questions, and sets the stage for future inquiry into perceptual learning from sentences.

One promising avenue for exploring this chapter's research questions lies in the reliability of evidence, which has been shown previously to be crucial for perceptual learning (Kraljic et al., 2008b,a). If ambiguous productions are accompanied by a video of the speaker holding a pen in their mouth, then no perceptual learning is observed (Kraljic et al., 2008b). Likewise, if listeners are first exposed to typical tokens, and then exposed to atypical tokens, no perceptual learning is observed. If the order is flipped (atypical tokens first), then perceptual learning effects are present (Kraljic et al., 2008b). If there is a linguistic context that conditions a greater variability, then modified tokens in those contexts will not cause perceptual learning (Kraljic et al., 2008a). However, the unlearned modification must lie within the range of variability conditioned by the context. For /s/ in /stɹ/ clusters, where /s/ becomes more /ʃ/ like due to coarticulation, a more /ʃ/-like modification will not be learned. Presumably, modifications that lay outside of the variability conditioned by the context will still be learned. Kraljic and colleagues argue from these studies that listeners will attribute variation to the context as much as possible, and only fall back to updating perceptual categories when no other explanation

is available.

Extended beyond single words, reliability can be thought of in terms of perceived carefulness of a word production. In an experimental setting, every stimulus is carefully curated by the experimenter. However, both in the laboratory and outside, words in isolation are produced longer and more clearly than their counterparts in full sentences. Words in isolated sentences are going to be produced less clearly than words in isolation (though not necessarily unintelligibly). Words in spontaneous conversation are likely to be the least clear, as seen in the "massive reduction" in the Buckeye corpus (Johnson, 2004; Dilts, 2013). All of these factors are dependent on aspects of the sentence (focus, clause type, etc.) or of the speech style, so words in casual conversation will be less clear than words in a formal presentation.

From a perception standpoint, the more clear an acoustic token, the more signal information is available to be processed. Clear tokens typically have longer durations, increased intensity, and more distinct formants (Krause and Braida, 2004). A listener would view tokens that were produced more clearly or with greater care as more reliable productions for that speaker. Listeners have been found to recognize careful and casual speech equally well, but signal information is used more in careful speech (Sumner et al., 2015). If we extend the argument by Kraljic and colleagues, it would predict that sentences should have less perceptual learning than words in isolation because (some of) the variability of a word's production in a sentence can be attributed to the fact that the item is in a sentence context. Additionally, given the propensity for acoustic reduction in high predictability contexts (Scarborough, 2010), words in predictable sentences would be even less reliable, leading to less perceptual learning.

This is not to say that sentences are ineffective in driving perceptual learning as compared to words in isolation. From the literature on perceptual learning of foreign accents, sentences are extremely useful in learning to perceive nonnatively accented speakers (Bradlow and Bent, 2008). For the purposes of learning an accent, sentences are probably better than words in isolation, as the greater context would allow for better identification of the words. Differences in perceptual learning from native and nonnative speakers can also be seen in the contradictory findings of Sumner (2011) and Kraljic et al. (2008b). Sumner (2011) found that

listeners could update their perceptual categories constantly over the course of the experiment. In contrast, Kraljic et al. (2008b) found that listeners adapted to the first instances of the category that they heard and did not use subsequent tokens. The nativeness of the exposure speaker differed in the two experiments. Constant adaptation was found for a nonnative speaker (Sumner, 2011) rather than a native speaker (Kraljic et al., 2008b). Listeners may be more biased toward typical native categories for native speakers, so that exposure to an initially typical category causes listeners to disregard the later atypical category as unreliable. Listeners can therefore leverage their previous experience with native speakers more readily. Listeners' previous experience does not as readily extend to nonnative speech, where interspeaker and intraspeaker variation is more prevalent, so constant adaptation and consistent token reliability would improve comprehension the most.

This dissertation largely adopts the predictive coding framework presented in Clark (2013) to account for perceptual learning. Reliability of sensory information is not directly addressed in Clark's exposition. The basic form of his model, however, predicts that increasing expectations should always increase error signals. In Egner et al. (2010, cited in Clark (2013)), participants were exposed to faces and non-face objects (i.e., houses) embedded in white noise on a computer screen (static). Participants who were told about the faces had equally-sized neuronal responses in the fusiform face area for face stimuli as for non-face stimuli. Participants who were not expecting to see faces showed neuronal responses in that area only for the face stimuli. The mismatch between expectation and the perceived signal generated an error signal of similar magnitude to the signal itself. If increased expectations result in increased error signals, perceptual learning should be largest for participants exposed to the modified category in higher predictability sentences. If smaller perceptual learning effects are observed for those participants, then reliability weighting or attribution of error signals would be necessary in the model.

To test these predictions, a novel exposure paradigm was used in place of a lexical decision task. In this paradigm, participants are presented with sentences auditorily. Following the sentence, two pictures appear on the screen: one matching the final word of the sentence and the other a distractor. Participants are instructed to indicate which picture corresponds to the final word of the sentence. Following

exposure, participants completed the same categorization task as those in Experiments 1 and 2. This experiment will validate lexical decision tasks for learning a single characteristic (/ʃ/-like /s/) in a context that is more closely resembling actual language use. At the same time, this experiment provides a link between lexically-guided perceptual learning and experiments that use sentential stimuli for learning non-native accents (Bradlow and Bent, 2008).

3.2 Methodology

3.2.1 Participants

A total of 137 participants from the UBC population completed the experiment and were compensated with either $10 CAD or course credit. The data from 39 nonnative speakers of English were excluded from the analyses. No participants reported speech or hearing disorders. This left data from 98 participants for analysis. Twenty additional native English speakers participated in a pretest to determine sentence predictability, and 10 other native English speakers participated in a picture naming pretest.

3.2.2 Materials

One hundred and twenty sentences were used as exposure materials. The set of sentences consisted of 40 critical sentences, 20 control sentences and 60 filler sentences. The critical sentences ended in one of 20 of the critical words in Experiments 1 and 2 that had an /s/ in the onset of the final syllable. The 20 control sentences ended in the 20 control items used in Experiments 1 and 2, and the 60 filler sentences ended in the 60 filler words in Experiments 1 and 2. Half of all sentences were written to be predictive of the final word, and the other half were written to be unpredictive of the final word. Unlike previous studies using sentence or semantic predictability (Kalikow et al., 1977), unpredictive sentences were written with a range of sentence structures. In all cases, the final words were plausible objects of lexical verbs and prepositions. The high and low predictability filler sentences can be found in Tables 3.1 and 3.2, respectively. The high and low predictability filler sentences with /ʃ/ words can be found in Tables 3.3 and 3.4, respectively. Fi-

nally, the high and low predictability critical sentences can be found in Tables 3.5 and 3.6, respectively. Aside from the sibilants in the critical and control words, the sentences contained no sibilants (/s z ʃ ʒ ʧ ʤ/). The same minimal pairs for phonetic categorization as in Experiments 1 and 2 were used.

Table 3.1: High predictability filler sentences.

Sentence	Word	Distractor
The oak tree grew from a tiny	acorn	pineapple
The radio in the car didn't work with a bent	antenna	towel
The clown made the girl an animal from a	balloon	pancake
Everyday the panda had to eat a lot of	bamboo	boomerang
The belt had an ornate	buckle	hamburger
The caterpillar came out of the cocoon a beautiful	butterfly	crayon
The hermit lived in a log	cabin	parade
They marked the date on the	calendar	antler
They rode to the pyramid on the back of a	camel	atom
Right before the plane took off, the captain called the flight crew from the	cockpit	doorknob
The woman threaded the bowtie through her	collar	ladybug
At the rodeo, the cattle were rounded up by the	cowboy	ripple
The baby rocked in her	cradle	telephone
The delivery man rang the	doorbell	firewood
He moved the wet laundry over to the	dryer	hydrant
The tiny rodent terrified the big, grey	elephant	pepper
The criminal wore a glove to not leave behind one	fingerprint	island
The cook needed one more clove of	garlic	wheelbarrow
Red paint in hand, the youth tagged the building with	graffiti	catapult
The watery dinner had to be poured out with a	ladle	lollipop
The man reheated the leftover dinner in the	microwave	ukulele
Every dinner plate came with a folded	napkin	toothpick
The ballroom had a grand	piano	dolphin
The woman tied her hair back in a	ponytail	airport
The adult frog developed from a	tadpole	bucket
The acting company performed in an old	theatre	earmuff
Her favorite burrito came wrapped in a flour	tortilla	falafel
The farm youth rode around on the	tractor	barbecue
The train went under a mountain through a	tunnel	wagon
The heavy rainfall could have been predicted by the	weatherman	robot

Sentences were recorded by the same male Vancouver English speaker used in Experiments 1 and 2. Critical sentences were recorded in pairs, with one normal

Table 3.2: Low predictability filler sentences.

Sentence	Word	Distractor
They clapped loudly for the	acrobat	pillow
The man liked to begin the day with an	apple	vampire
Wearily, the woman built up her	campfire	bagel
He looked forward to freely available	candy	donkey
The couple never agreed on the	cutlery	butter
He took pride in the renovated	darkroom	candle
They were enthralled by the	diamond	kiwi
While he lay on the ground, the boy played with a	feather	broccoli
To get any farther, they definitely needed a good	goalie	waterfall
He didn't know how to get to the	gondola	honey
They were a little frightened to board the	helicopter	cannon
The woman needed to borrow a	ladder	flagpole
He had to track down and get help from the	librarian	tornado
They had a good view of the	lightning	coffee
Toward the end, they were running low on	lumber	anvil
They liked how it looked on the	mannequin	parrot
On the way, they liked walking through the	meadow	cupcake
The couple were looking forward to buying a	minivan	tugboat
He finally made it to the	motel	armadillo
They went out for a quick bite to eat after the	movie	volleyball
He really liked the look of the	mural	monocle
After a long night, he devoured the whole	omelet	hummingbird
On the field trip, they learned all about the	painter	pulley
The boy cried when they took away the	popcorn	puppet
The irate woman yelled at the	referee	propeller
When they were called, the group moved to the	table	helmet
He didn't know about a problem with the	teapot	crowbar
He had to remember to pick up the	tire	parakeet
Every day he dreaded the late afternoon	traffic	rowboat
The woman kept a lookout for the	umbrella	catamaran

production and then a production of the same sentence with the /s/ in the final word replaced with an /ʃ/. The speaker was instructed to produce both sentences with comparable speech rate, speech style, and prosody.

As in Experiments 1 and 2, the critical items were morphed together into an 11-step continuum using STRAIGHT (Kawahara et al., 2008); only the final word in each sentence was morphed. The preceding words were the synthesized versions of the sentence with the correct /s/ production to minimize artifacts of the morphing

Table 3.3: High predictability sentences with /ʃ/ words.

Sentence	Word	Distractor
The bidding became frantic for the final item in the	auction	accordion
While waiting in line at the new bank, the woman read their introductory	brochure	blueberry
The woman only got a dime back after paying the	cashier	laptop
He could only kneel for a little while without a plump	cushion	forklift
Lava flowed down the volcano after the violent	eruption	pumpkin
The bear awoke from her winter	hibernation	violin
After jumping out of the plane, the woman opened her	parachute	camera
The doctor took the time to look in on every	patient	crocodile
While down with the flu, the woman invariably carried a clean	tissue	whirlpool
The opera-goer found her row with the help of an	usher	doormat

Table 3.4: Low predictability sentences with /ʃ/ words.

Sentence	Word	Distractor
The woman couldn't wait to fill up the	bookshelf	muffin
The whole family travelled for an hour to the	coronation	waffle
He did not look forward to the	handshake	raccoon
He gave a wide berth to the	machine	kitten
They dragged their feet on the way to the	mansion	treadmill
He had a hard time with	meditation	beekeeper
They were deeply worried about the	militia	peanut
He went on and on about the	milkshake	elbow
For winter break, he wanted to go to the	ocean	iguana
He could finally get a new	windshield	koala

algorithm. The control and filler items were also processed and resynthesized to ensure consistent quality. The ambiguous point selection was based on the pretest performed for Experiment 1 and 2 exposure items. The ambiguous steps of the continua chosen corresponded to the 50% cross over point in Experiment 1.

Acoustic distances between exposure tokens, categorization tokens, and their original productions were multidimensionally scaled. In Figure 3.1, the original productions are separated again by the first dimension, which corresponds to the centroid frequency of the sibilant. The categorization tokens are predictably in between the original productions and offset in the second dimension due to their

Table 3.5: High predictability sentences with target /s/ words.

Sentence	Word	Distractor
At the carnival the girl rode a unicorn around the	carousel	pirate
A deep moat protected the old	castle	martini
The encore from the pop duo perfected the whole	concert	earplug
From the bakery he got a flaky, buttery	croissant	windmill
After her world trip, the traveller had a little money leftover in every local	currency	elevator
When the computer locked up, he couldn't move the	cursor	clover
The lady returned the bow with a formal	curtsy	gavel
The critic raved about the ballet and the lead	dancer	cricket
Long ago, a comet hit the earth, killing every big	dinosaur	bandana
Water poured into the bath tub from the	faucet	doughtnut
After millennia, the bone in the riverbed turned into a	fossil	menorah
The name 'Milky Way' can perfectly depict our	galaxy	kayak
We no longer worry about the plague due to modern	medicine	cucumber
In the heated aerial battle, neither pilot could lock on with a	missile	cookie
Rain fell every day in India during the	monsoon	gargoyle
The man wrote on the paper with a graphite	pencil	trombone
The woman got an over-the-counter drug at her local	pharmacy	kettle
From the cap of the new grad hung a golden	tassel	guitar
The New Yorker flagged down a	taxi	ribbon
The traffic cop alerted the driver by blowing her	whistle	ravioli

different position in the word. The exposure tokens for Experiment 3 fit in between the original productions and the categorization tokens.

Sentences were pairs with two pictures apiece. Pictures of 200 words, with 100 pictures for the final word of the sentences and 100 for distractors, were selected in two steps. First, a research assistant selected five images from a Google image search of the word, and then a single image representing that word was selected from amongst the five by me. To ensure consistent behavior in E-Prime (Psychology Software Tools, 2012), pictures were resized to fit within a 400x400 area with a resolution of 72x72 DPI and converted to bitmap format. Additionally, any transparent backgrounds in the pictures were converted to plain white backgrounds.

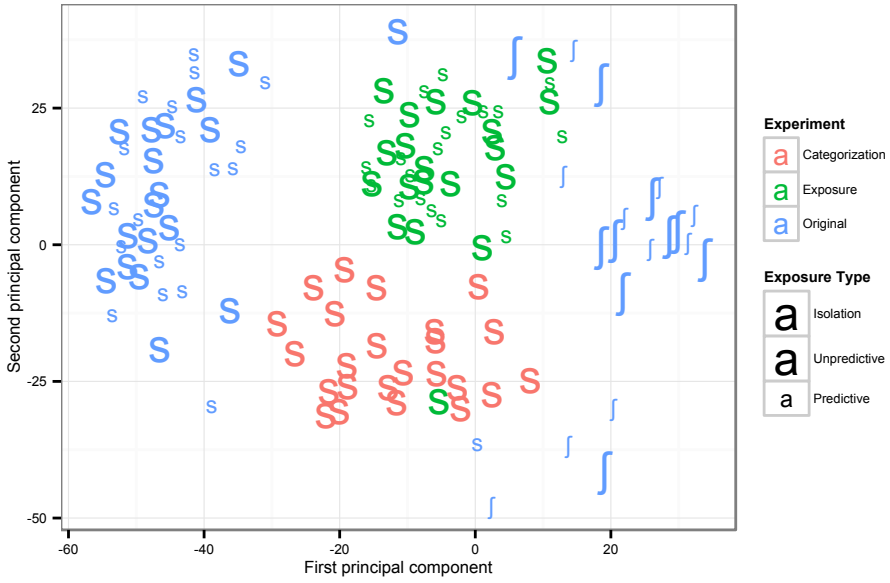
Table 3.6: Low predictability sentences with target /s/ words.

Sentence	Word	Distractor
They got back in line for the	carousel	pirate
He dreaded the long walk to the	castle	martini
He prepared night and day for the	concert	earplug
The man had a craving for a	croissant	windmill
They weren't worried about the different	currency	elevator
The man could never find the	cursor	clover
The girl didn't want to make a	curtsy	gavel
The boy wanted to become a better	dancer	cricket
The boy really wanted to ride the	dinosaur	bandana
The woman hoped to get a working	faucet	doughtnut
No one knew where to find the	fossil	menorah
The man talked at length about the	galaxy	kayak
With that GPA, they could have a career in	medicine	cucumber
The boy wanted to build a toy	missile	cookie
On their picnic, they avoided the	monsoon	gargoyle
The woman looked frantically for her	pencil	trombone
The woman loved her work at the	pharmacy	kettle
He worried about the color of the	tassel	guitar
The woman had no luck getting a	taxi	ribbon
The boy ran away when he heard the	whistle	ravioli

3.2.3 Pretest

The same twenty participants that completed the lexical decision continua pre-test also completed a sentence predictability task before the phonetic categorization task described in Experiment 1. Participants were compensated with $10 CAD for both tasks, and were native North American English speakers with no reported speech, language or hearing disorders. In this task, participants were presented with the 120 exposure sentences with the final target word removed. Participants were instructed to type in the word that came to mind when reading the fragment, and to enter any additional words that came to mind that would also complete the sentence. There was no time limit for entry and participants were shown an example with the fragment "The boat sailed across the..." and the possible completions "bay, ocean, lake, river". Responses were collected in E-Prime (Psychology Software Tools, 2012), and were sanitized by removing miscellaneous keystrokes, spell checking, and standardizing variant spellings and plural forms.

Figure 3.1: Multidimensional scaling of the acoustic distances between the sibilants of original productions, categorization tokens and the exposure tokens in Experiment 3. Note that the only Isolation tokens are the Categorization tokens.



From the sanitized data, responses were coded as either 0 if the target word was not present or 1 if it was. For each sentence, the target response rate was calculated by averaging responses from all participants. The target response rate was 0.49 (range 0-0.95) for predictive sentences and 0.03 (range 0-0.45) for unpredictive sentences. Predictive sentences that had target response rates of 0.2 or less were rewritten. The predictive sentences for *auction*, *brochure*, *carousel*, *cashier*, *cockpit*, *concert*, *cowboy*, *currency*, *cursor*, *cushion*, *dryer*, *graffiti*, and *missile* were rewritten to remove any syntactic or semantic ambiguities. For instance, a common completion for the predictive sentence "The youth tagged the wall with..." was "spray paint" rather than "graffiti". To promote the likelihood of "graffiti", the sentence was changed to "Red paint in hand, the youth tagged the wall with...", which would eliminate "spray paint" as a possible completion.

Five volunteers participated in another pretest to determine how suitable the pictures were at representing their associated word. All participants were native speakers of North American English, with reported corrected-to-normal vision. Participants were presented with a single image in the middle of the screen. Their task was to type the word that first came to mind, and any other words that described the picture equally well. There was no time limit and presentation of the pictures was self-paced. Responses were sanitized as above.

Pictures were replaced if 20% or less of the participants (1 of 5) responded with the target word and the responses were semantically unrelated to the target word. Five pictures were replaced, *toothpick* and *falafel* with clearer pictures and *ukulele*, *earmuff* and *earplug* were replaced with *rollerblader*, *anchor* and *bedroom*. All five replacements were for distractor words.

3.2.4 Experiment design

Participants were assigned to one of four groups from a 2x2 between-subject factorial design. The first factor was whether the word containing the ambiguous sibilant was predictable from the preceding words or not (Predictability: Predictive versus Unpredictive). All participants were therefore exposed to a consistent 100 stimulus sentences with identical control and filler items for all participants. The second factor was whether participants were given additional instructions about the sibilant or not (Attention: Attention versus No Attention). Participants in the Attention condition received additional instructions that the speaker's "s" sounds were sometimes ambiguous, and to listen carefully to ensure correct responses.

3.2.5 Procedure

As in Experiments 1 and 2, participants completed an exposure task and a categorization task in E-Prime (Psychology Software Tools, 2012). For the exposure task, participants heard a sentence via headphones for each trial. Immediately following the auditory presentation, they were presented with two pictures on the screen. Their task was to select the picture on the screen that corresponded to the final word in the sentence they heard. The order was pseudorandom with the same constraints described in Experiment 1. Half of the matching pictures were selected via one

button and half via the other.

Each trial proceeded as follows. A blank screen was presented for 250 ms. Immediately following, a sentence was presented auditorily. Following the auditory stimulus, two pictures and their respective buttons appeared on the screen. For example, a sentence ending in "dog" would show a picture of a dog and "1" on one side of the screen, and a picture of a banana and "5" on the other side of the screen. Participants had up to 3000 ms to respond which picture matched the final word in the sentence. Feedback as to whether a response was detected was shown for 500 ms before the next trial began. Participants were given a self-paced break after 50 trials.

Following the exposure task, participants completed the same categorization task described in Experiments 1 and 2.

Participants were given oral instructions explaining both tasks at the beginning of the experiment to remove experimenter interaction between exposure and categorization. Written instructions were presented to participants at the beginning of each task as well. The instructions for the exposure task given to participants assigned to an Attention condition included explicit reference to the modified sibilants. Participants were told that "this speaker's 's' sound is sometimes ambiguous" and instructed to "listen carefully so as to choose the correct response."

3.2.6 Analysis

Response data and factors were transformed and analyzed in the same way as in Experiment 1 and 2.

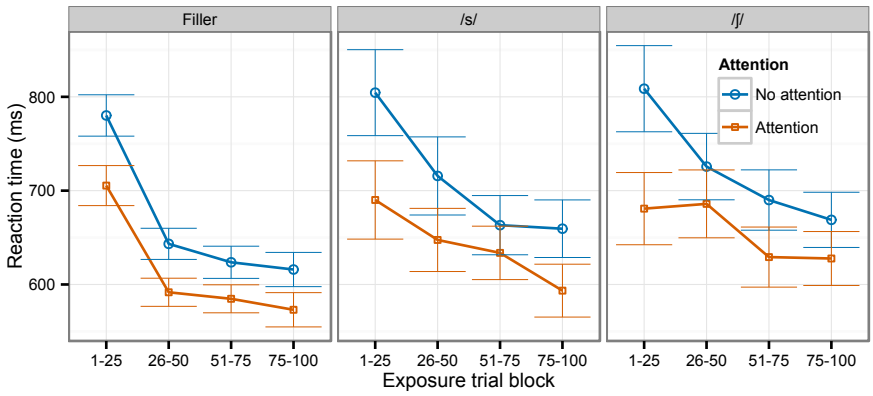
3.3 Results

3.3.1 Exposure

Performance in the task was high, with accuracy near ceiling across all subjects (mean accuracy = 99.5%, sd = 0.8%). Due to these ceiling effects, a logistic mixed-effects model of accuracy was not constructed. A linear mixed effects model for logarithmically-transformed reaction time was constructed with a similar structure as in Experiments 1 and 2. Fixed effects were Trial (0-100), Trial Type (Filler, /s/,

and /ʃ/), Attention (No Attention and Attention), Predictability (Unpredictive and Predictive), and their interactions. By-Subject and by-Word random effect structure was as maximal as permitted by the data, with by-Subject random slopes for Trial, Trial Type, Predictability, and their interactions and by-Word random slopes for Attention, Predictability, and their interaction. Trial Type was coded using treatment (dummy) coding, with Filler as the reference level. Deviance contrast coding was used for Predictability (Unpredictive = 0.5, Predictive = -0.5) and Attention (No attention = 0.5, Attention = -0.5).

Figure 3.2: Within-subject mean reaction time in the exposure phase of Experiment 3, separated out by Trial Type (Filler, /s/, and /ʃ/). Error bars represent 95% confidence intervals.



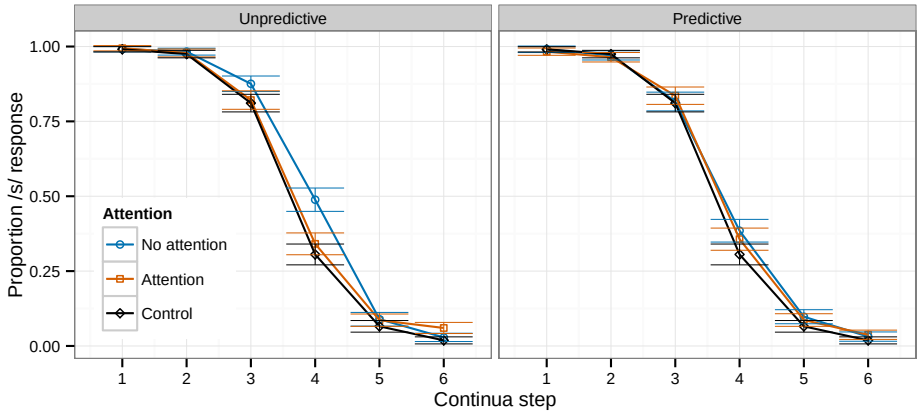
A significant effect was found for Trial ($\beta = -0.20, SE = 0.01, t = -11.0$), indicating that reaction time became faster over the course of the experiment. There was a significant effect for Trial Type of /ʃ/ versus Filler ($\beta = 0.19, SE = 0.09, t = 2.1$), but not for /s/ versus Filler ($\beta = 0.11, SE = 0.09, t = 1.3$), suggesting that words with /ʃ/ in them were responded to more slowly than filler words or those with a modified /s/ in them. There was a significant interaction between Trial and Trial Type of /s/ versus Filler ($\beta = 0.05, SE = 0.02, t = 2.4$) and between Trial and Trial Type of /ʃ/ versus Filler ($\beta = 0.05, SE = 0.02, t = 2.9$), indicating that reaction time for words with /s/ or /ʃ/ in them did not become as fast across the experiment

as those for filler words. These results are shown in Figure 3.2. Note that the y-axis has a different scale than that used in Experiments 1 and 2 for reaction times. Participants were faster in general in this task than in the lexical decision task. Responses to predictable sentences (mean = 669 ms, SD = 321 ms) were not significantly faster than responses to unpredictable sentences (mean = 646 ms, SD = 299 ms), suggesting that performance was at floor.

3.3.2 Categorization

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. A logistic mixed effects model was constructed with Subject and Continua as random effects and continua Step as random slopes, with 0 coded as a /ʃ/ response and 1 as a /s/ response. Fixed effects for the model were Step, Exposure Type, Attention, and their interactions, with deviance coding used for contrasts for Exposure Type (Unpredictive = 0.5, Predictive = -0.5) and Attention (No attention = 0.5, Attention = -0.5).

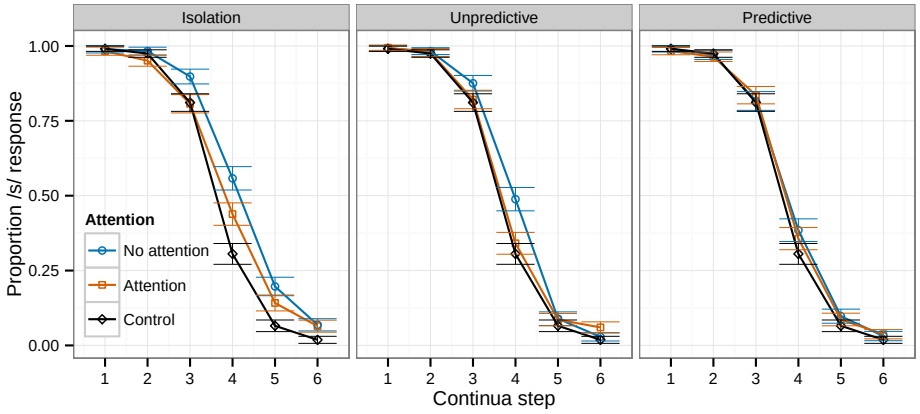
Figure 3.3: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 3. Error bars represent 95% confidence intervals.



As in the previous experiments, there was a significant effect of the intercept

($\beta = 0.52, SE = 0.20, z = 2.6, p < 0.01$) and of Step ($\beta = -2.49, SE = 0.19, z = -12.7, p < 0.01$). Exposure Type ($\beta = 0.23, SE = 0.23, z = 0.97, p = 0.33$), Attention ($\beta = 0.30, SE = 0.21, z = 1.4, p = 0.15$), and their interaction ($\beta = 0.38, SE = 0.44, z = 0.9, p = 0.39$) are all not significant, despite the visible differences in Figure 3.3. In Figure 3.3, there appears to be a similar interaction pattern as was seen for Experiment 1 (Figure 2.6). Participants in the different attention conditions for Unpredictive exposure appear to differ in Step 4. However, the lack of significance suggests that this may be less reliable or more localized to Step 4 than in Experiment 1.

Figure 3.4: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 3 and the word-medial condition of Experiment 1. Error bars represent 95% confidence intervals.



As an additional comparison, the data from this experiment was combined with the subset of participants in Experiment 1 who were exposed to the same set of words (the word-medial condition). Exposure Type was recoded as a three-level factor, using treatment (dummy) contrast coding, with the Experiment 1 exposure (Isolation) as the reference level. An identically specified logistic mixed effects model was fit to this data set as to the initial data. In this model, there was a significant effect of Attention ($\beta = 0.74, SE = 0.32, z = 2.2, p = 0.02$), such that participants in Attention conditions were less likely to categorize the continua steps

as /s/. Exposure Type had a marginal effect of Predictive compared to Isolation ($\beta = -0.43, SE = 0.23, z = -1.9, p = 0.05$), indicating that participants in the Predictive condition were less likely to categorize the continua as /s/ overall as compared to participants from Experiment 2. Step interacted with both Unpredictive as compared to Isolation ($\beta = -0.42, SE = 0.17, z = -2.4, p = 0.01$) and Predictive as compared to Isolation ($\beta = -0.32, SE = 0.14, z = -2.2, p = 0.02$). These interactions indicate that the categorization functions for sentential stimuli had a steeper cross-over than the Isolation. As shown in Figure 3.4, the endpoints (Steps 5 and 6) for the sentential conditions are wholly overlapping with the control categorization for those steps. While participants in the Unpredictive condition showed a shifted category boundary, the perceptual learning affected less of the continua than for participants in the Isolation condition.

An additional model was run with the reference level for Exposure Type as Predictive to check whether participants in the Predictive condition showed perceptual learning effects at all. In the model with Predictive as reference, the intercept is no longer significant ($\beta = 0.44, SE = 0.28, z = 1.5, p = 0.12$), indicating perceptual learning was not robustly present in participants in the Predictive condition. The difference between the Predictive condition and the Isolation condition remains ($\beta = 0.77, SE = 0.32, z = 2.4, p = 0.01$), and, as above, the difference between Predictive and Unpredictive is not significant ($\beta = 0.42, SE = 0.31, z = 1.3, p = 0.17$). These results indicate participants in the Predictive condition showed no perceptual learning effects, and participants in the Unpredictive condition were in between Predictive and Isolation, but not significantly different from either. Increasing the statistical power might separate the conditions further.

3.4 Discussion

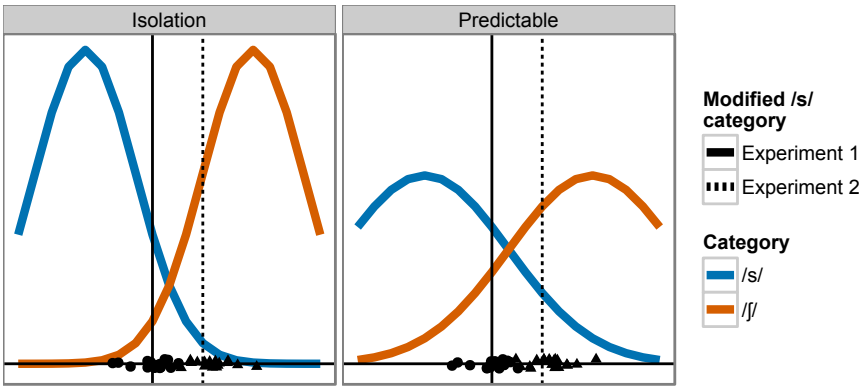
The key finding of the current experiment is that modified categories embedded in words in meaningful sentences produce less perceptual learning than words in isolation. In fact, participants exposed to a modified category only in predictive words had a similar boundary as those in the control experiment who had no exposure to a modified /s/ category. This pattern of results aligns the most with the extension to Kraljic and colleagues' argument that perceptual learning is a last resort. If there

is any way to attribute the acoustic atypicality to either linguistic or other sources of variation, no perceptual learning occurs (Kraljic et al., 2008b,a). In the current experiment, semantic predictability may be a linguistic source to which variation can be attributed. Semantic predictability shows effects that are similar to a more local source like consonant cluster coarticulation.

The prediction of a simple predictive coding model (Clark, 2013) was not borne out. Rather than increased expectations enhancing error signals, the conditions with increased expectations showed no perceptual learning at all. How can we reconcile then the predictive coding model and the findings of the current experiment? One way, certainly, is to incorporate the reliability argument of Kraljic and colleagues. Bayesian approaches capture uncertainty quite well, so the unreliable tokens, such as those in the high predictability sentences, would have greater uncertainty associated with them. Another possibility is that perceptual learning did occur, but it was not generalized to the test items. Participants could have learned from their exposure how the speaker produces /s/ in high predictability contexts, but the context of words in isolation was too different from the exposure context. Put another way, the participants could have learned how the speaker reduces his /s/ category, but not how the speaker normally produces it.

However, if semantic predictability functions in a similar way as consonant cluster coarticulation, listeners would not show perceptual learning effects even if they were tested on a continuum in a high predictability sentence. In Kraljic et al. (2008a), listeners exposed to an ambiguous /s/ in the context of /stɹ/ and then tested on a continuum from /astɹi/ to /aʃtɹi/ showed no perceptual learning effects. Participants who were exposed to ambiguous /s/ intervocalically showed perceptual learning on both /asi/-/aʃi/ and /astɹi/-/aʃtɹi/ continua. There was no exposure-specificity effect, so participants did not even learn that the speaker produces a more /ʃ/-like /s/ in that context. Any abstract encoding process accounts for and removes the variability associated with the context, leaving the unmodified perceptual category. A similar pattern is likely to be seen with high predictability exposure. Importantly, the speaker's durations for the target /s/ words did not different across predictability conditions (Predictable /s/ words: mean = 0.53 s, SD = 0.06 s; Unpredictable /s/ words: mean = 0.53 s, SD = 0.07 s). Any effect of predictability is more likely from listener perception than speaker production in this experiment.

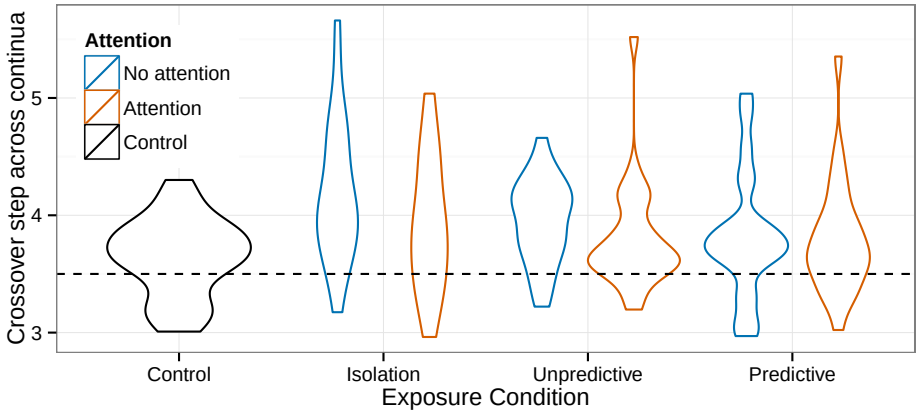
Figure 3.5: Schema of category relaxation in predictable sentences. The solid vertical line represents the mean of the modified category similar to the one used for Experiment 1, and a dashed vertical line represents the mean of the Experiment 2 modified category. A more atypical category, as was used in Experiment 2, has a higher probability of being categorized as /s/ in predictable sentences than in isolation.



One question raised by this finding is whether perceptual learning is possible at all in high predictability sentences. If the range of acceptable variation for all categories is expanded (schematized in Figure 3.5), the modified category would have fallen closer to the expected mean in predictable sentences compared to isolation. In terms of error propagation, the modified categories used here may not have generated enough errors to learn from. Presenting listeners with a more atypical category should then cause more perceptual learning in this case. In Figure 3.5, the atypical category from Experiment 2 would have a higher likelihood of being categorized as /s/ in predictable sentences than in isolation. If increasing the atypicality in predictable sentences did in fact result in increased perceptual learning, it would suggest that perceptual learning is maximized in a particular range. Tokens too close to the expected mean are too typical to learn from, and tokens too far from the expected mean are too unreliable. However, if listeners simply ignore atypical sounds in highly predictable words, then increasing the atypicality of the category

(i.e. using the ambiguity threshold from Experiment 2) would not increase perceptual learning. If that were the case, listeners might not even be sensitive to replacing the /s/ with another sound category entirely (i.e. /f/) in a comprehension-oriented task (but see Samuel, 1981).

Figure 3.6: Distribution of cross-over points for each participant across comparable exposure tokens in Experiments 1 and 3. Larger bulges represent more subjects located at that point in the distribution. The dashed line represents the mean step of the continua. Large bulges around the dashed line for Control, Unpredictive and Predictive conditions indicate that many speakers did not change their category boundaries, compared to the Isolation conditions.



As a final point in this discussion, the distribution of individuals' perceptual learning effects differs in shape as compared to Experiment 1. Figure 3.6 shows the distribution of cross-over points of each subject in Experiment 3 and participants in the condition of Experiment 1 that used the same exposure words. Cross-over points are where along the continua perception switches from primarily /s/ to primarily /ʃ/, and higher cross-over points are indicative of greater perceptual learning. Participants exposed to the modified category in sentences show more consistency (larger bulges in the violin plots) than those exposed in isolated words. In the Isolation conditions, participants follow a fairly wide, even distribution. In contrast,

participants in the sentence conditions a more tightly clustered either around the normalized cross over point or a step above, suggesting potentially discrete groups in the distribution.

One possible reason for these more discrete groups may relate to cognitive load. Under lower cognitive load conditions, participants in perception-oriented tasks show greater perceptual sensitivity. In this experiment, the task is comprehension-oriented, so lower cognitive load could have distributed cognitive resources either to the comprehension task or to aspects of the signal. Participants with better attention-switching control might devote those resources to perception, while those with worse attention-switching control might not, which may be the cause of the findings in Scharenborg and Janse (2013). Future research should quantify participants' attention-switching abilities and other individual differences that may play a role in explaining these findings.

Chapter 4

Discussion and conclusions

This dissertation set out to examine the influence of listener attentional sets on perceptual learning. Perceptual learning is a phenomenon common to many fields involved in cognitive science. How perceptual learning generalizes to new contexts, however, is quite different across paradigms. Perceptual learning in psychophysics is the process of a perceiver aligning their senses to the world. Perceptual learning in speech perception is the process by which perceivers align their perceptual system to an interlocutor to facilitate understanding.

I argue that perceptual learning as a mechanism is shared between linguistic and non-linguistic domains. However, psychophysics paradigms employ primarily perception-oriented attentional sets, while speech perception paradigms employ both perception-oriented and comprehension-oriented attentional sets. Perception-oriented attentional sets in all domains lead to less generalized learning. Conversely, comprehension-oriented attentional sets lead to more generalized learning. The first two experiments of this dissertation implement a standard lexically-guided perceptual learning paradigm – a lexical decision task – but with manipulations promoting perception-oriented attentional sets. Even in a comprehension-oriented lexical decision task, promoting more perception-oriented attentional sets leads to less generalized learning. These results provide a crucial link between fully comprehension-oriented perceptual learning in the lexically-guided paradigm and fully perception-oriented perceptual learning in visually-guided paradigms. The remainder of this chapter first summarizes the results of the dissertation as they re-

late to specificity and generalization in the perceptual learning literature. The four manipulations used to promote the different attentional sets are then examined, followed finally by implications for models of cognition and psycholinguistics.

4.1 Specificity and generalization in perceptual learning

The results of this dissertation speak to the dichotomy between specificity and generalization found in the perceptual learning literature. In Experiment 1, participants had larger perceptual learning effects when they were exposed to ambiguous sounds later in the words rather than at the beginning of words (e.g. *carousel* versus *cement*). And yet, the testing continua consisted of stimuli with the sibilant at the beginnings of words, which are more similar to the exposure tokens beginning with the ambiguous sound. Exposure that matched the word position of the categorization (word-initial) showed no greater perceptual learning effects than word-medial exposure. Perceptual learning, therefore, occurred at a level of abstraction that is usually not assumed in perceptual learning studies. Most lexically-guided perceptual learning studies attempt to make the exposure tokens and the categorization similar – and in some cases, the same – in order to maximize exposure-specificity effects. In this dissertation, listeners generalized from stimuli with large degrees of coarticulation (i.e., in the middle of the word) to stimuli without as much coarticulation. In some cases, the perceptual learning effect was largest in precisely the cases where coarticulatory effects differed from exposure to test. One aspect that was not tested in the current studies is exposure-specificity at the level of the item. Perhaps a more perception-oriented attentional set would show greater perceptual learning on the specific exposure items.

The effect of attentional set manipulations in Experiments 1 and 2 suggest that when listeners adopt more perceptually-oriented attentional sets, even within tasks that are oriented toward comprehension, generalization of perceptual learning to new forms is inhibited. Lexically-guided perceptual learning is more likely to be expanded to new contexts than visually-guided perceptual learning (Norris et al., 2003; Kraljic et al., 2008a; Reinisch et al., 2014, but see Mitterer et al., 2013). Visually-guided and psychophysical perceptual learning paradigms often have highly repetitive stimuli with little variation. Both of these aspects add to

the monotony of the task and the likelihood of perception-oriented attentional sets (Cutler et al., 1987).

Lexically-guided perceptual learning, on the other hand, requires very few instances to affect the perceptual system. The standard number is around 20 ambiguous tokens within 200 trials, but as few as 10 ambiguous tokens have been shown to have comparable effects (Kraljic et al., 2008b). A consequence of the proposed attention mechanism is that it nicely captures the different number of stimuli needed for perceptual learning across comprehension-oriented and perception-oriented tasks. Tokens heard under comprehension-oriented attentional sets should have a relatively large effect on the perceptual system as compared to tokens heard under a perception-oriented attentional set. A single token updating a more abstract representation will generalize more than many repetitions updating fine-grained episodic representations. From this, we could predict that word endorsement rate and category boundary shift would be less linearly correlated the more comprehension-oriented participants are. This prediction is borne out by the lack of correlation between word endorsement rate and cross-over point in Experiment 1 in the Word-final/No Attention condition. This condition is predicted to have the most comprehension-oriented attentional set of the conditions, and here is the only instance in Experiment 1 where a significant correlation between word endorsement rate and cross-over point is not present. Participants in this condition have relatively high cross-over points that do not depend as much on the sheer number of tokens endorsed.

4.2 Effect of increased linguistic expectations

The conditions of Experiment 1 that are most similar to previous lexically-guided perceptual learning paradigms are those with no explicit instructions about the /s/ category. In these conditions, increasing linguistic expectations through lexical bias resulted in larger perceptual learning effects. I argue that the increased perceptual learning is due to increased maintenance of comprehension-oriented attentional sets by participants in the Word-medial condition. The participants exposed to a modified /s/ category at the beginnings of words would be more likely to have their attention drawn to the atypicality of the modified /s/ category. There are two

potential scenarios for how this would have affected participants. In the first, "normal" word processing would proceed with the perception of the modified /s/ as part of comprehending the word, but the attentional set would not change. In the second, processing the word would trigger an attentional set change that would get reinforced for each new modified /s/ encountered. The experiments in this dissertation do not definitively answer which scenario is more likely, and it could be that different participants fall into different scenarios. However, when participants were told about the ambiguity of the /s/, they do not behave any differently if the /s/ is word-initial or word-medial. This similarity of behavioral patterning suggests the second scenario is more likely, and more perception-oriented attentional sets were adopted as a result of exposure to words beginning with a modified /s/ category.

Increasing linguistic expectations through semantic predictability did not increase perceptual learning. In fact, there was a trend towards unpredictive sentences increasing perceptual learning. Semantic predictability has previously been shown to affect perception-oriented tasks in a similar way as lexical bias (Connine, 1987; Borsky et al., 1998). In Experiment 3, however, participants exposed to the modified /s/ category in high predictability sentences showed no perceptual learning effects at all. While the Isolation condition (Word-medial condition in Experiment 1) was not significantly different from the Unpredictive condition of Experiment 3, there was a trend toward reduced perceptual learning when the modified sound category was embedded in a sentential context in general. The lack of a perceptual learning effect from high predictability exposure sentences is reminiscent of studies that find no perceptual learning when a modified /s/ category is embedded in a /stɹ/ cluster that conditions that variation (Kraljic et al., 2008a). In both cases, the modified category is embedded in a context that conditions increased variability. However, there is a difference between the consonant cluster context and the semantic predictability context. In the consonant cluster, there is a straightforward coarticulatory reason for /s/ to surface as more /ʃ/-like in /stɹ/ clusters, with the /s/ produced more in a postalveolar position due to the upcoming /ɹ/. For semantic predictability, there is no particular reason why a /s/ should surface more /ʃ/ like in high predictability sentences. If high semantic predictability can be the attributed cause of /s/ surfacing as more /ʃ/-like, it seems reasonable that the range of acceptable productions for all categories would be expanded (as

schematized in Figure 3.5 of Chapter 3).

Perceptual learning of nonnative accents is possible through hearing sentences of varying predictability (Bradlow and Bent, 2008, and others). However, the phonetic variability involved in those tasks reaches far beyond the sibilant modified here. The speaker producing the sentences in this dissertation is a native English speaker of the local dialect. Even with the synthesis applied to the sound files, he is more intelligible than the speakers in studies involving nonnative accents. The ease of comprehension of the speaker in this study might actually inhibit perceptual learning in sentences, because listeners can leverage so much of their perceptual experience with other speakers of the local dialect.

On the flip side, how nonnative listeners perceptually adapt to speech that varies in predictability is an interesting question as well. Nonnative listeners do not benefit from high semantic predictability as much as native listeners (Mayo et al., 1997). This tends to result in less accuracy for transcribing speech in noise. As the sentences presented here did not include noise, the lessened benefit from semantic predictability might manifest differently. If high predictability sentences are not as predictable for those listeners, they may show perceptual learning effects more similar to unpredictable sentences.

4.3 Attentional control of perceptual learning

The findings of Experiment 1 support the hypothesis that comprehension-oriented attentional sets produce larger perceptual learning effects than perception-oriented attentional sets. Although all participants showed perceptual learning effects, those exposed to the ambiguous sound with increased lexical bias only showed larger perceptual learning effects when the instructions about the speaker's ambiguous sound were withheld. Attention on the ambiguous sound equalized the perceptual learning effects across lexical bias. However, in Experiment 2, there is no such effect of attention. This suggests that ambiguous sounds farther away from the canonical production induce a more perception-oriented attentional set regardless of explicit instructions.

One question raised by the current results is whether perception-oriented attentional sets always result in decreased perceptual learning. The instructions used to

focus the listener's attentional set framed the ambiguity in a negative way, with listeners being cautioned to listen carefully to ensure they made the correct decision. If the attention were directed to the ambiguous sound by framing the ambiguity in a positive way (i.e., that the ambiguous "s" is from a non-native accent or a speech disorder), would we still see the same pattern of results? The current mechanism would predict that attention of any kind to signal properties would block the propagation of errors, reducing perceptual learning. This prediction will be tested in future work.

Attention's role in perceptual learning may extend to the realm of sociolinguistics. In sociolinguistics, there are three categories of linguistic variables: indicators, markers, and stereotypes (Labov, 1972). Of these, stereotypes are the most known to speakers of the dialect and speakers of other dialects. If attention to perception inhibits perceptual learning, then perceptual learning of these stereotyped linguistic variables would be inhibited. For instance, New Zealand English has undergone several vowel shifts as compared to other varieties of English, but these shifted vowels differ in salience depending on the listener's dialect (Bell, 2015). For Australian English listeners, the STRUT vowel is salient (*fish and chips* as more *fush and chups*). For North American English listeners, the DRESS vowel is more salient (*Bret* heard as *Brit*). These two populations of listeners are predicted to perceptually adapt to these vowel changes differently. North American English listeners should adapt to STRUT more than Australian Englsih listeners, and vice versa for DRESS. Given the scale from indicators to markers to stereotypes is ordered in terms of speaker (or listener) awareness, the role of attention proposed in this dissertation would predict progressively less perceptual learning as awareness increases. Salient social variants (i.e. r-lessness) have also been found to not be encoded as robustly as canonical productions (Sumner and Samuel, 2009). Are less salient social variants learned easier in general?

4.4 Category atypicality

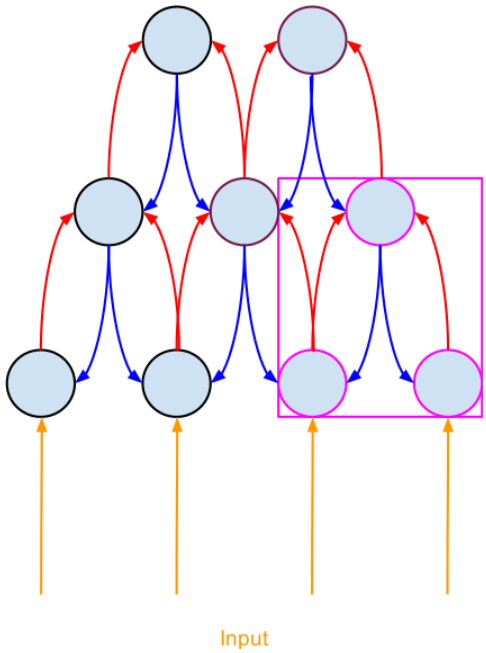
In Experiment 2, there was no effect of explicit instructions or lexical bias on perceptual learning, with a stable perceptual learning effect present for all listeners. There are two potential, non-exclusive explanations for the lack of effects. As

stated above, the increased distance to the canonical production drew the listener's attention to the ambiguous productions, resulting in a perception-oriented attentional set. The second potential explanation is that the productions farther from canonical produce a weaker effect on the updating of a listener's categories, as predicted from the neo-generative model in Pierrehumbert (2002). This explanation is supported in part by the weaker correlation between word endorsement rate and cross-over point found in Experiment 2, and the findings of Sumner (2011) where the highest rates of perceptual learning were found when the categories began more typical and gradually became less typical over the course of exposure. This explanation could be tested straightforwardly by implementing the same gradual shift paradigm used in Sumner (2011) with the manipulations used in this dissertation.

An interesting extension to the current findings would be to observe the perceptual learning effects in a cognitive load paradigm. Speech perception under cognitive load has been shown to have greater reliance on lexical information due to weaker initial encoding of the signal (Mattys and Wiget, 2011). Following exposure to a modified ambiguous category, we might expect to see less perceptual learning if the exposure was accompanied by high cognitive load. Scharenborg et al. (2014), however, found that hearing loss of older participants did not significantly influence their perceptual learning. Therefore it may be that perceptual learning would not fluctuate across cognitive loads. Higher cognitive loads, however, might allow for more atypical ambiguous stimuli to be learned, due to the increased reliance on lexical information during initial encoding.

It is important to bear in mind that what is typical in one context is not necessarily typical in another. The methodology employed for Experiment 3 assumed that expected variation for the category /s/ would be common across all experiments. However, it may be that the perfectly ambiguous /s/ category in Experiment 3 was within the range of variation in high predictability sentences. In this case, had the category atypicality been more like that of Experiment 2, we may have actually seen more of an effect, perhaps back to the level of Experiment 1 (as schematized in Figure 3.5 of Chapter 3).

Figure 4.1: A schema for predictive coding under a perception-oriented attentional set. Attention is represented by the pink box, where gain is enhanced for detection, but error signal propagation is limited to lower levels of sensory representation where the expectations must be updated. This is represented by the lack of pink nodes outside the attention box. As before, blue errors represent expectations, red arrows represent error signals, and yellow represents the sensory input.

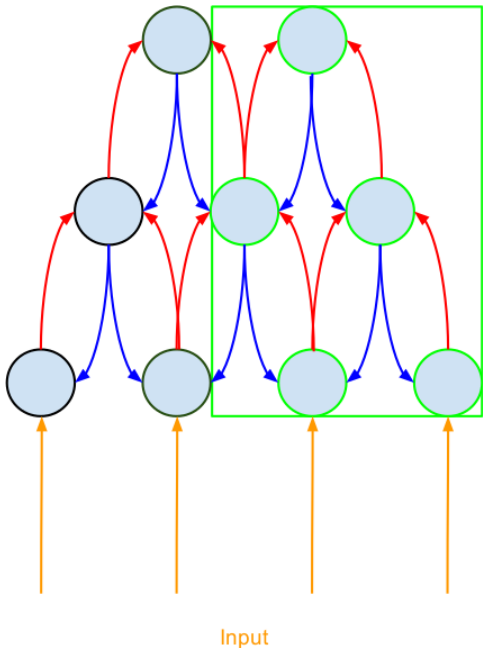


4.5 Implications for cognitive models

The model that this dissertation adopts is based off of the predictive coding framework (Clark, 2013). In this model, expectations about incoming signal are fed from higher levels of representation to lower ones. The mismatch between actual perceived signal and the expectations is then propagated back to the higher levels as an error signal. Future expectations are modified based on the error signal. This framework captures the basics of perceptual learning, and a similar computational framework has been used to model visually-guided perceptual learning tasks (Kleinschmidt and Jaeger, 2011). However, the attentional mechanism in the

predictive coding framework does not work well for some instances of visual attention (Block and Siegel, 2013) or for the current results. I propose a new attentional mechanism for predictive coding, one in which attention inhibits error propagation beyond the level to which attention is directed. Figures 4.1 and 4.2 show schemas reproduced from Chapter 1 for perception-oriented and comprehension-oriented attentional sets, respectively. Such a mechanism explains both the previous findings and the current results.

Figure 4.2: A schema for predictive coding under a comprehension-oriented attentional set. Attention is represented by the green box, where it is oriented to higher, more abstract levels of sensory representation. Error signals are able to propagate farther and update more than just the fine grained low level sensory representations. As before, blue errors represent expectations, red arrows represent error signals, and yellow represents the sensory input.



The predictive coding framework advanced here has implications for other psycholinguistic research outside of perceptual learning. Recent innovations in speech

perception models have emphasized the role of episodic memory traces (Goldinger, 1996; Pierrehumbert, 2001). That is, listeners encode more phonetic detail than is strictly necessary for linguistic comprehension, and can process previously heard tokens of a word type faster than novel tokens of that word type. Theodore et al. (2015) argue that attention during encoding can emphasize abstract (i.e. lexical) information at the expense of episodic (i.e., talker) information or vice versa. Such a proposal is similar to that put forth in this dissertation, but encoding in a predictive coding framework would be updating of predictions. The lack of an explicit memory trace mechanism in the predictive coding framework may be a weakness concerning cognition as a whole, but I would argue that it still accounts for the speech perception data. The principle motivation for episodic memory traces was originally to account for behavioral data that showed sensitivity to fine details of previous stimuli (Goldinger, 1996). However, Sumner and Kataoka (2013) highlights recent findings of recognition equivalence but memory inequality between frequent forms and idealized, infrequent forms. For instance, the word *flute* is generally pronounced with an unreleased /t/ in North American English, but it is also produced less frequently with a fully released /t/ (the idealized form) or with a glottal stop. All pronunciation variants are recognized equally well in short term processing tasks (accuracy and reaction time). However, infrequent, idealized pronunciations are remembered better in long term recall tasks. Sumner and colleagues propose an alternate route to linguistic encoding, which they term socioacoustic encoding. Hierarchical respresentations in the predictive coding framework can account for this data without appealing to episodic memory traces. The socioacoustic encoding would be a speaker-based hierarchical representation, with abstracted gender and accent representations.

While most of the discussion here has concerned representations of speech sounds, predictive coding representations are not solely limited to linguistic objects. In fact, one of the key findings of perceptual learning is that it is largely dependent on the speaker. In these cases, perceptual learning is not updating just the distribution for what is expected of a speech sound, but also what is expected for that speaker. Perceptual learning from a group of speakers that share a trait (i.e., the same non-native accent) facilitates the creation (or perhaps simply the identification) of a more abstract category for that group of speakers, enhancing

intelligibility on future novel speakers (Bradlow and Bent, 2008).

The predictive coding framework can be applied to speech production as well, and has particular applicability to sound change. When a person produces speech, they are also perceiving it and compensating for any deviations from their predictions (Hickok et al., 2011). In addition to hierarchical representations for what a person's own speech should sound like, there could also be social representations that act on it. If a speaker identifies with a particular speech group, then their predictions for their own speech should align with what they expect other memebers of the speech group to produce. One way of thinking about speech style in production (e.g. reading, interview, casual conversation, etc.) is in terms of attention to speech production (Labov, 1997). As attention to speech production increases, speech group markers (i.e., non-rhoticity in New York City) become less prevelant (Labov, 1997). Perhaps attention plays an inhibitory role on abtract social representations in speech production. In terms of sound change, an individual would change their speech both when they associate a particular trait with a speech group and consider themselves a member of that group.

Recent work on a historical vowel change shift in New Zealand English proposed that low frequency words led the shift (Hay et al., 2015). The mechanism they propose to account for this data is one where tokens that are difficult to comprehend are less likely to be encoded. Low frequency words are particularly affected because they are likely to be interpreted as higher frequency neighbors and nonwords. The experiments of this dissertation contained a similar situation at an individual speaker level. Participants were more likely to not recognize words containing a modified /s/ category as real English words than the filler words. In Experiment 1, the amount to which a participant's boundary shifted was – in general – correlated with the amount of /s/ words recognized as words. To the extent that difficult to comprehend words are treated as nonwords, these findings reinforce the findings of Hay et al. (2015).

Outside of psycholinguistcs, this dissertation suggests testable predictions for perceptual learning in the visual domain using visual illusions. In the Kanizsa illusion, for instance, three Pac-man like objects are arranged to give the illusion of three circles overlaid by a triangle (Kanizsa, 1976). Perception of this illusion requires more abstract representations that are not in the signal, much like the objects

of comprehension as defined in this dissertation. The proposed mechanism for attention would predict that perceptual learning involving visual illusions should be more general and less exposure specific. In the Kanizsa case, perceivers would perceptually learn characteristics of the abstract triangles and circles instead of the Pac-man shapes. Visual illusions allow perceivers to better organize complex scenes in short-term and working memory (Vandenbroucke et al., 2012). Similar to these illusions, words and higher linguistic structures allow better organization of complex auditory signals. Drawing attention to either the circles and triangles of the illusion or to the Pac-man symbols should induce attentional sets similar to comprehension-oriented and perception-oriented ones proposed in this dissertation, with similar effects on the generalization of perceptual learning.

I have argued that attentional sets, particularly within the predictive coding framework, are crucial to the generalization of perceptual learning to new contexts. Recently proposed models of speech perception treat linguistic representations as a balance of both more abstract elements and more fine-detailed elements (Theodore et al., 2015) and also incorporate aspects of social representations (Szakay, 2012; Sumner and Kataoka, 2013). Both of these trends are easy to incorporate into a predictive coding framework. Such a model accounts for the findings of this thesis and those of the larger psycholinguistic literature.

4.6 Conclusion

This dissertation investigated the influences of attention and linguistic salience on perceptual learning in speech perception. Perceptual learning was modulated by the attentional set of the listener. Perception-oriented attentional sets were induced through increasing the salience of the modified category, either by reducing the lexical bias, increasing the typicality, or giving explicit instructions. In all these instances, participants showed robust perceptual learning effects, but smaller effects than participants not biased towards perception-oriented attentional sets. Exposure to modified categories in predictive sentences resulted in no perceptual learning effects, potentially due to the attribution of the modified sound category to reduced speech clarity. These results support a greater role of attention than previously assumed in predictive coding frameworks, such as the proposed propagation-

blocking mechanism. Finally, these results suggest that the degree to which listeners perceptually adapt to a new speaker is under their control to the same degree as attentional set adoption. However, given the robust perceptual learning effects found across experiments, perceptual learning is a largely automatic process when variation cannot be attributed to contextual factors.

Bibliography

Ahissar, M. and Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12):5718–22. → pages 19

Bacon, W. F. and Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, 55(5):485–496. → pages 2, 3

Baker, A., Archangeli, D., and Mielke, J. (2011). Variability in american english s-retraction suggests a solution to the actuation problem. *Language variation and change*, 23(03):347–374. → pages 25

Bell, A. (2015). The indexical cycle and the making of social meaning in language: Why everybody needs good neighbours sociolinguistically. Colloquium at University of British Columbia. → pages 84

Bertelson, P., Vroomen, J., and De Gelder, B. (2003). Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect. *Psychological Science*, 14(6):592–597. → pages 9

Block, N. and Siegel, S. (2013). Attention and perceptual adaptation. *Behavioral and Brain Sciences*, 36(3):205–6. → pages 21, 56, 87

Borsky, S., Tuller, B., and Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America*, 103(5 Pt 1):2670–2676. → pages 16, 58, 82

Bradlow, A. R. and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4):2339–2349. → pages 16, 58

Bradlow, A. R. and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2):707–729. → pages 1, 10, 60, 62, 83, 89

Brysbaert, M. and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4):977–990. → pages 31

Clare, E. (2014). Applying phonological knowledge to phonetic accommodation. In *Poster presented at the 14th Conference on Laboratory Phonology*, Tachikawa, Tokyo. → pages 15

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204. → pages ix, 9, 10, 14, 20, 22, 43, 55, 56, 61, 75, 86

Clopper, C. G. and Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, 124(3):1682–1688. → pages 16, 58

Connine, C. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 538:527–538. → pages 16, 58, 82

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19(2):141–177. → pages 3, 13, 14, 18, 81

Dilts, P. C. (2013). *Modelling Phonetic Reduction in a Corpus of Spoken English Using Random Forests and Mixed-effects Regression*. PhD thesis, University of Alberta. → pages 60

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, 30(49):16601–16608. → pages 61

Eisner, F. and McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2):224–238. → pages 11, 44

Eriksen, B. A. and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1):143–149. → pages 19

Fallon, M., Trehub, S. E., and Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America*, 111(5 Pt 1):2242–2249. → pages 16, 58

Finn, A. S., Lee, T., Kraus, A., and Kam, C. L. H. (2014). When it hurts (and helps) to try: The role of effort in language learning. → pages 17

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110–125. → pages 12

Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin*, 50(6):401–431. → pages 2, 7, 17

Gilbert, C., Sigman, M., and Crist, R. (2001). The neural basis of perceptual learning. *Neuron*, 31:681–697. → pages 11, 22, 26

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. → pages 4, 88

Gow, D. W. and Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2):344–359. → pages 5, 14

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., and LaShell, P. (2015). Tracking word frequency effects through 130years of sound change. *Cognition*, 139:83–91. → pages 89

Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, 69(3):407–422. → pages 89

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54. Citeseer. → pages 60

Kalikow, D., Stevens, K., and Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61(5). → pages 15, 16, 58, 62

Kanizsa, G. (1976). Subjective contours. *Scientific American*, 234(4):48–52. → pages 89

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-straight: A temporally stable power spectral representation

for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3933–3936. → pages x, xi, 31, 36, 48, 64

Kleber, F., Harrington, J., and Reubold, U. (2012). The Relationship between the Perception and Production of Coarticulation during a Sound Change in Progress. *Language and Speech*, 55(3):383–405. → pages 53

Kleinschmidt, D. and Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the 2nd ACL Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics. → pages 9, 86

Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008a). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, 107(1):54–81. → pages 15, 25, 59, 75, 80, 82

Kraljic, T. and Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2):141–78. → pages 11, 15, 44

Kraljic, T. and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1):1–15. → pages 11

Kraljic, T., Samuel, A. G., and Brennan, S. E. (2008b). First impressions and last resorts: how listeners adjust to speaker variability. *Psychological Science*, 19(4):332–8. → pages 15, 24, 25, 55, 59, 60, 61, 75, 81

Krause, J. C. and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1):362–378. → pages 60

Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6):1668–1679. → pages 1

Labov, W. (1972). Sociolinguistic Patterns. *Language*, 2(4):344. → pages 84

Labov, W. (1997). The Social Stratification of (r) in New York City Department Stores. In Coupland, N., editor, *Sociolinguistics: A Reader*, pages 168–178. St. Martin's Press. → pages 89

Leber, A. B. and Egeth, H. E. (2006). Attention on autopilot: Past experience and attentional set. *Visual Cognition*, 14(4-8):565–583. → pages 18, 20

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177. → pages 17

Lieberman, P. (1963). Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech. *Language and Speech*, 6(3):172 –187. → pages 16

Ling, S. and Carrasco, M. (2006). When sustained attention impairs perception. *Nature neuroscience*, 9(10):1243–1245. → pages 20

Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1):29–63. → pages 5, 14

Mattys, S. L. and Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160. → pages 13, 17, 18, 58, 85

Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40(3):686–693. → pages 16, 58, 83

McAuliffe, M. (2015). python-acoustic-similarity. Available from https://github.com/mmcauliffe/python-acoustic-similarity. → pages 34

McLennan, C. T., Luce, P. a., and Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):539–53. → pages 10

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, 122(2):145–163. → pages 34

Mirman, D., McClelland, J. L., Holt, L. L., and Magnuson, J. S. (2008). Effects of Attention on the Strength of Lexical Influences on Speech Perception: Behavioral Experiments and Computational Mechanisms. *Cognitive Science*, 32(2):398–417. → pages 5, 13

Mitterer, H., Scharenborg, O., and McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2):356–361. → pages 11, 80

Norris, D. and Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, 43(6):541–550. → pages 18

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2):204–238. → pages 2, 6, 7, 11, 15, 19, 22, 26, 44, 55, 80

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology 7*. → pages 55, 85

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. and Hopper, B., editors, *Frequency and the emergence of linguistic structure*, pages 137–158. John Benjamins, Amsterdam. → pages 24, 88

Pitt, M. and Szostak, C. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, (April 2013):37–41. → pages 3, 5, 6, 13, 14, 18, 27, 28

Pitt, M. A. and Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language*, 29(5):611–632. → pages 18

Pitt, M. A. and Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4):699–725. → pages 14

Pitt, M. A. and Samuel, A. G. (2006). Word length and lexical activation: longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5):1120–1135. → pages 5, 13, 27

Psychology Software Tools, I. (2012). E-Prime. → pages 32, 37, 66, 67, 69

Reinisch, E. and Holt, L. L. (2013). Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2):539–555. → pages 11

Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):75–86. → pages 2, 11, 22, 37, 39, 44, 55

Reinisch, E., Wozny, D. R., Mitterer, H., and Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45:91–105. → pages 2, 11, 22, 26, 54, 55, 80

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and motor skills*, 8(3):271–276. → pages 19

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, 8(2):101–105. → pages 17

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4):474–494. → pages 14, 16, 25, 58, 77

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In *Laboratory phonology 10*, pages 575–604. → pages 16, 58, 60

Scharenborg, O. and Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception & Psychophysics*, 75(3):525–36. → pages 20, 51, 78

Scharenborg, O., Weber, A., and Janse, E. (2014). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception, & Psychophysics*, pages 1–15. → pages 19, 20, 55, 85

Shankweiler, D., Strange, W., and Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. → pages 1

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119(1):131–136. → pages 22, 24, 28, 60, 61, 85

Sumner, M. and Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, 134(6). → pages 1, 88, 90

Sumner, M., McGowan, K., D'Onofrio, A., and Pratt, T. (2015). Casual speech is more sensitive to top-down information than careful speech. In *Oral presentation at the Linguistic Society of America 2015 Annual Meeting*. → pages 60

Sumner, M. and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4):487–501. → pages 84

Szakay, A. (2012). *The effect of dialect on bilingual lexical processing and representation*. PhD thesis, University of British Columbia. → pages 90

Theodore, R. M., Blumstein, S. E., and Luthra, S. (2015). Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, Perception, & Psychophysics*, pages 1–11. → pages 4, 88, 90

Toro, J. M., Sinnett, S., and Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2). → pages 17

Vandenbroucke, A. R. E., Sligte, I. G., Fahrenfort, J. J., Ambroziak, K. B., and Lamme, V. A. F. (2012). Non-Attended Representations are Perceptual Rather than Unconscious in Nature. *PLoS ONE*, 7(11). → pages 90

Vroomen, J., van Linden, S., de Gelder, B., and Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3):572–577. → pages 7, 9, 24, 55

Watanabe, T., Náñez, J. E., and Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413(6858):844–848. → pages 17

Wolfe, J. M. and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501. → pages 18

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706):72–75. → pages 21, 56