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 ARTS
(Linguistics)

The University Of British Columbia
(Vancouver)

April 2015

Abstract

This dissertation presents three experiments investigating perceptual learning in speech perception under different attention and salience conditions. Listeners across all three experiments were exposed to a speaker's /s/ category that had been modified to sound more like /ʃ/. All listeners showed a perceptual learning effect on a categorization task following exposure, categorizing more of continua from /s/ to /ʃ/ as /s/ than a control group that completed only the categorization task, replicating previous lexically-guided perceptual learning results (e.g., Norris et al., 2003; Reinisch et al., 2013). However, the magnitude of perceptual learning effects differed depending on the linguistic context of the modified /s/ category and the listener's attentional set. The linguistic context of modified categories has previously been shown to affect a wide range of psycholinguistic tasks, such as phoneme identification (Pitt & Samuel, 2006; Borsky et al, 1998) and phoneme restoration (Samuel, 1981). The attentional set of the listener has also been shown to affect how tolerant they are to modified categories (Pitt & Szostak, 2012). This dissertation replicates these studies in the context of lexically-guided perceptual learning, showing that these factors that influence the online processing of speech also influence the longer term adaptation to speakers. In Experiment 1, listeners exposed

to the modified category in the middle of words showed larger perceptual learning effects than those exposed to the category at the beginnings of words. However, this difference in magnitude was only observed when listeners were not told that the speaker's /s/ sounds would be ambiguous. When listeners were given these instructions, the difference between word-initial and word-medial exposure was not present. In Experiment 2, the exposure conditions were the same as Experiment 1, but the modified /s/ category was more atypical, sounding more like /ʃ/ than /s/ rather than halfway in between. In this experiment, exposure conditions had no effect, and all participants showed the same size of perceptual learning effect. In Experiment 3, listeners were exposed to the same modified /s/ category as Experiment 1 in word-medial position only, but the words were embedded at the end of sentences. These sentences differed in how strongly they conditioned the final word, with one group of listeners hearing the modified category only in sentences that were predictive of the target word, and the other group only hearing the category in sentences that were not predictive of the target word. The same pattern as Experiment 2 is present for Experiment 3, with no significant differences in exposure conditions. These results are framed in a predictive coding framework (Clark, 2013), which is a hierarchical Bayesian model of the brain. Predictions from higher levels are sent to lower ones, and mismatches between these predictions and the lower level's input cause an error signal to propagate back to the higher levels. The expectations for future stimuli are then tuned based on this error signal. The model captures perceptual learning results well as a whole, but the mechanism for attention in the framework is a simple gain based one, where attention gives increased weight to the error signals. Such a mechanism would predict that attention would result in increased perceptual learning, but that prediction is not supported

by results of this dissertation's experiments. Instead, increased attention to low-level perception reduces the amount of perceptual learning observed. Increased attention to perception can be induced in different ways either through explicit instructions, linguistic prominence, or the salience of the atypicality of the category. I propose a mechanism of attention in the predictive coding framework that inhibits propagation of error signals beyond the level to which attention is oriented. Attention focused on perception will inhibit the propagation of error signals to higher levels, limiting the updating of expectations at higher levels and leading to reduced perceptual learning. Perception-oriented tasks, such as psychophysical visual perceptual learning or visually-guided perceptual learning in speech perception, show a lack of generalization, with effects limited to exposure items (Gilbert et al., 2001; Reinisch et al., 2014), but comprehension-oriented tasks, such as lexically-guided perceptual learning, are more likely to show generalization to new items (Norris et al., 2003; Reinisch et al., 2013). The propagation-inhibiting mechanism for attention can account for the degrees of generalization reported across the literature.

Revision: May 11, 2015

Preface

At University of British Columbia (UBC), a preface may be required. Be sure to check the Graduate and Postdoctoral Studies (GPS) guidelines as they may have specific content to be included.

Table of Contents

List of Tables

List of Figures

Glossary

This glossary uses the handy `acroynym` package to automatically maintain the glossary. It uses the package's `printonlyused` option to include only those acronyms explicitly referenced in the LaTeX source.

GPS Graduate and Postdoctoral Studies

Acknowledgments

Thank those people who helped you.

Molly!

Kathleen!

530 class!

Jobie!

Jamie!

Michelle!

Don’t forget your parents or loved ones.

Mom!

Dad!

Laura!

You may wish to acknowledge your funding sources.

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 in categorization studies (Shankweiler et al., 1977; Kuhl, 1979) and as recognition equivalence in word recognition tasks (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. There is also variability on the part of the listener. They may be distracted by concurrent demands or they may be focused on different aspects of the speech. If the speaker has a peculiar trait, a listener may listen specifically for that trait, to the detriment of overall understanding. Perceptual learning is typically framed in terms of speaker variation. This dissertation examines the effect of listener variation on perceptual learning.

In the speech perception literature perceptual learning refers to the updating of distributions corresponding to a sound category, such as /s/ or /ʃ/, following exposure to speech with a modification to the sound category's distribution (Norris et al., 2003). Exposure to a speaker affects a listener's perceptual system, even after only a few tokens (Vroomen et al., 2007; Kraljic et al., 2008b) or within the span of a single utterance (Ladefoged and Broadbent, 1957). The changes to the perceptual system are commonly specific to that particular speaker (cf Eisner and McQueen, 2005; Kraljic and Samuel, 2007). Perceptual learning has also been shown for groups of speakers with common characteristics, such as accent (Bradlow and Bent, 2008).

Generalization to novel contexts has been a locus of investigation within the perceptual learning literature. Across most studies, exposure to modified sound categories generalizes well 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. On the other hand, generalization appears to be more limited when perceptual learning is induced through visually-guided paradigms in which listeners are exposed to a modified category through matching it to an unambiguous video signal. The perceptual learning exhibited from these visually-guided experiments is only found to influence 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).

Why, then, does lexically-guided perceptual learning produce such generaliza-

tion? 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 elicted, 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 attentional set is the feature-detection set – a focused set limited to the target's defining feature. If there are multiple, redundant targets or many distractor singletons, then the singleton-dectection 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 requires less effort than feature search. Even if participants are given instructions targeting a relevant feature, they will default to singleton detection unless it is made ineffective through stimuli 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 comprehen-

sion 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, due to their focus on stimuli that are devoid of linguistic meaning.

Comprehension of a word necessarily involves perception.

The 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. To test this 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, and one is instruction-based, and the fourth is stimuli-based. The rest of this section is devoted to an overview of these manipulations and their motivations.

The first linguistic manipulation that affects 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. 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, word-initial exposure is predicted to promote a more perception-oriented attentional set, and word-medial exposure is predicted to promote a more comprehension-oriented attentional set. Experiments 1 and 2 use this manipulation in Chapter 2, and further background is given in Section 1.2.1 of this chapter.

The second linguistic manipulation employed is the context that a word appears in. In Experiments 1 and 2, participants are exposed to the modified sound category in words in isolation, 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. Using 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 groups 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 stimuli-based manipulation is the 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. A more atypical category is predicted to promote a perception-oriented attentional set than a more typical category because its atypicality is predicted to be more perceptually salient.

Perceptual salience is relevant in this dissertation only in so far as it promotes a perception-oriented attentional set. Salience is a widely-used and poorly-defined term across literatures. For the purposes of this dissertation I adopt the following definition: an element is salient if it is unpredictable from context and/or easily distinguishable from other possible elements. The sound category that is being learned in this dissertation already has salient signal properties (e.g., in the form of high frequency, relatively high amplitude, aperiodic noise). Therefore, increasing salience of the modified /s/ category is a function of embedding it a linguistic position with little conditioning context. Increasing the acoustic distance of the modified category to the typical distribution of /s/, and in turn increasing its distinguishability, will also increase the salience of the category and promote a more perception-oriented attentional set.

The results of these experiments will be contextualized in the existing perceptual learning literature and in models of perceptual learning. Section 1.1 reviews the perceptual learning literature in greater depth, as well as conceptual models that have been proposed for it (Clark, 2013; Kleinschmidt and Jaeger, 2011). Broadly

speaking, perceptual learning is well accounted for by a Bayesian model, where differences between expectations and observed input cause an error signal which updates future expectations, leading to perceptual learning. The results of the experiments in this dissertation support a more nuanced attention mechanism, however, than that proposed in Clark (2013). The mechanism for attention proposed by Clark is gain-based such that increased attention increases the weight of error signals Thus, increased attention should lead to increased perceptual learning. Generalization of perceptual learning, however, is dependent on the task and the attentional set being employed. An expanded attention mechanism is therefore proposed to unify the results of perceptual learning studies across the psychophysics and speech perception literatures, encompassing the results of the current experiments. The proposed mechanism inhibits error propagation beyond the level where attention is directed. In a perception-oriented attentional set, perceptual learning is hypothesized to occur at the initial perception where encoding is more fine-grained, inhibiting generalization. Comprehension-oriented attentional sets are predicted to allow for greater error propagation and abstraction, leading to greater levels of generalization.

The structure of the thesis is as follows. This chapter reviews the recent literature on perceptual learning in speech perception (Section 1.1), as well as literature on linguistic expectations (Section 1.2), attention (Section 1.3), and category typicality (Section 1.4) as they relate to the three experiments of this dissertation. Chapter 2 will detail 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 second experiment using a more atypical /s/ category that is more /ʃ/-like than /s/-like. Chapter 3 details an experiment using a novel perceptual learning paradigm that manipulates semantic predictability to increase the linguistic expectations during exposure. Finally, Chapter 4 summarizes the results and places them in a theoretical framework. 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 manipulate the two primary attentional sets.

1.1 Perceptual learning

Perceptual learning is a well established phenomenon in the psychology and 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 (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 training that assumes attention on the task, but does not require reinforcement, correction, or reward. This definition of perceptual learning corresponds more to what is termed "selective adaptation" in the speech perception literature rather than what is termed "perceptual learning", "perceptual adaptation" or "perceptual recalibration." In speech perception, selective adaptation is the phenomenon where listeners that are exposed repeatedly to a narrow distribution of a sound category narrow their own perceptual category. This results in a change in variance of the category, but not of the mean of the category (along some acoustic-phonetic dimension) (Eimas et al., 1973; Samuel, 1986; Vroomen et al., 2007).

Perceptual learning or recalibration in the speech perception literature is a more broad updating of perceptual categories either in mean or variance (Norris et al., 2003; Vroomen et al., 2007).

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 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 behaviour, 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 linguistic 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/.

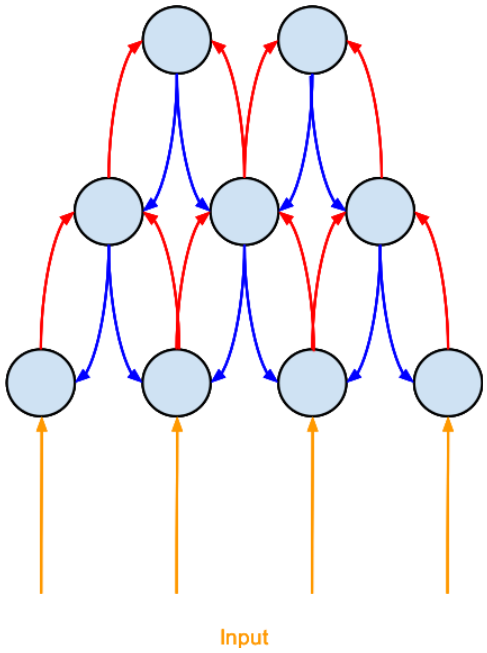
van Linden and Vroomen (2007) compared the perceptual recalibration effects from the visually-guided perceptual learning paradigm (Bertelson et al., 2003) to the standard lexically-guided perceptual learning paradigm (Norris et al., 2003). Speech-read recalibration and perceptual learning effects had comparable sizes, lasted equally as long, were enhanced when presented with a contrasting sound, and were both unaffected by periods of silence between exposure and categorization. The perceptual learning effects for both lexically- and visually-guided paradigms did not last through prolonged testing. However, in Kraljic and Samuel (2005) and Eisner and McQueen (2006), lexically-guided perceptual learning effects persisted through intermediate tasks, as long as the task did not involve any contradictory evidence of the trait learned.

Visually-guided perceptual learning in speech perception has been modeled using a Bayesian belief updating framework (Kleinschmidt and Jaeger, 2011). A more general Bayesian framework for perception and action in the brain is the predictive coding model (Clark, 2013), schematized in Figure 1.1. In Bayesian belief updating, the model categorizes the incoming stimuli based on multimodal cues, and then updates the distribution to reflect that categorization. This updated con-

ditional distribution is then used for future categorizations in an iterative process. Kleinschmidt and Jaeger (2011) effectively model the results of the behavioural 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. A similar, but more broad, framework is that of predictive coding (Clark, 2013). 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 then is the result of modifying expectations to match learned input.

Perceptual learning in the psychophysics literature has shown a large degree of exposure-specificity, where observers only show learning effects on the same or very similar stimuli as those they were trained on. As such, perceptual learning has been argued to reside or affect the early sensory pathways, where stimuli are represented with the greatest detail (Gilbert et al., 2001). Perceptual learning in speech perception 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 across exposure and testing (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 would be expected from a purely psychophysical standpoint. The testing stimuli are in many ways quite different from the exposure stimuli, with participants trained on multisyllabic words ending in an ambiguous sound and tested on monosyllabic words (Reinisch et al., 2013) and nonwords

Figure 1.1: Schema of the predictive coding framework (Clark, 2013). Nodes are sensory representations that become more abstract the higher they are in the hierarchy. Blue arrow represent expectations, red arrows are error signals, and yellow is the actual sensory input.



(Norris et al., 2003; Kraljic and Samuel, 2005), though 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, attentional and 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 experiments in this dissertation use the lexically-guided perceptual learning paradigm, which uses tasks oriented towards comprehension, so generalization is to be expected in general, but the more perception-oriented the attentional set, the less perceptual learning should be observed. In terms of a Bayesian framework with error propagation, a more perception-oriented attentional set may keep error propagation more local, resulting in the exposure-specificity seen more in the psychophysics literature. A more comprehension-oriented attentional set would propagate errors farther upward in the hierarchy of expectations. In both cases, errors would propagate to where attention is focused, but larger updating associated with farther error propagation would lead to larger observed perceptual learning effects. 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, sententially-guided perceptual learning paradigm is used to further promote use of comprehension-oriented attentional sets.

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 particular meaningful word rather than a nonsense word. For instance, given a continuum from a nonword like *dask* to word like *task* that differs only in the initial sound, listeners are more likely to interpret any step along the continuum as the word endpoint rather than the nonword endpoint 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. Given that ambiguous productions must be associated with a word to induce lexically-guided perceptual learning (Norris et al., 2003), differing degrees of lexical bias could lead to differing degrees of perceptual learning, a prediction which is tested in Chapter 2. The stronger the lexical bias, the stronger the link will be between the ambiguous sound and the general sound category.

Lexical bias 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 one attentional set or the other. Increasing cognitive load has been found to increase lexical bias effects as well. Performing a hard task concurrent to a phoneme categorization task results in a greater weighting

of lexical information, due to impoverished auditory encoding (Mattys and Wiget, 2011). Both of these tasks are perception-oriented as well, and the prevalence of effects from comprehension-oriented attentional sets speaks to its prominent role in ordinary language use.

Aspects of the stimuli items themselves have a large impact on the size of lexical bias effects. 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 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). They argue that longer words exert stronger lexical bias due to both more conditioning information present in longer words and the greater lexical competition for shorter words.

Different positions within words of a given length 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*. In the

experiments and meta-analysis of phoneme identification results presented in Pitt and Samuel (1993), they found that for monosyllabic word frames token-final targets produce more robust lexical bias effects than token-initial targets. Considering word length and position in the word, lexical bias appears to be 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. 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.

Sounds at the beginnings of words show more fragile lexical bias effects than those later in the word. Sounds at the beginning of words require more accurate perception than sounds later in words, because they are used to narrow the set of potential words and no particular expectations are available before the word be-

gins. As such, perception-oriented attentional sets are favored in the processing of word-initial sounds. Lexically-guided perceptual learning typically maximizes the lexical bias of the modified category by placing it at the end of the word (Norris et al., 2003), but perceptual learning has been found when the ambiguous stimuli is embedded earlier in the word (e.g., the onset of the final syllable (Kraljic and Samuel, 2005; Kraljic et al., 2008b,a) or the onset of the first syllable (Clare, 2014)). However, in light of the findings in Pitt and Szostak (2012), we expect less lexical bias when the ambiguous sounds occur earlier in the word. Word-initial modified sound categories are predicted to have lower word endorsement rates and smaller perceptual learning effect sizes. Experiments 1 and 2 in Chapter 2 implement this manipulation.

1.2.2 Semantic predictability

The second type of linguistic expectation manipulation used in this dissertation is semantic predictability (Kalikow et al., 1977). Sentences are semantically predictable when they contain words prior to the final word that 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.

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, high predictability sentences are generally more intelligible. Sentences that form a semantically coherent 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 bilingual listeners, but not late bilingual listeners (Mayo et al., 1997). Highly predictable sentences are more intelligible to native listeners in noise, even when signal enhancements are not made, though non-native listeners require both signal enhancements and high predictability together 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 cue 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 stimuli 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, greater phonetic encoding is available (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 that 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. 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 on perceptual learning in the psychophysics literature, indeed Gibson (1953) identifies it as the sole prerequisite to perceptual learning. As an example, 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 this visual domain is limited to the task and stimuli on which a given participant was trained.

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 attentional sets used here. For instance, in a visual search task, colour, 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 where primary attention is on detecting words from nonwords in a lexical decision task, and a focused attentional set is where instructions direct participants' attention to a potentially misleading sound. Attentional set selection is primarily affected by the instructions and the stimuli for the task. Attentional sets 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 temporal 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 cog-

nitive load, such as performing a more difficult concurrent task, listeners show increased lexical bias effects (Mattys and Wiget, 2011). Variation in speech 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.

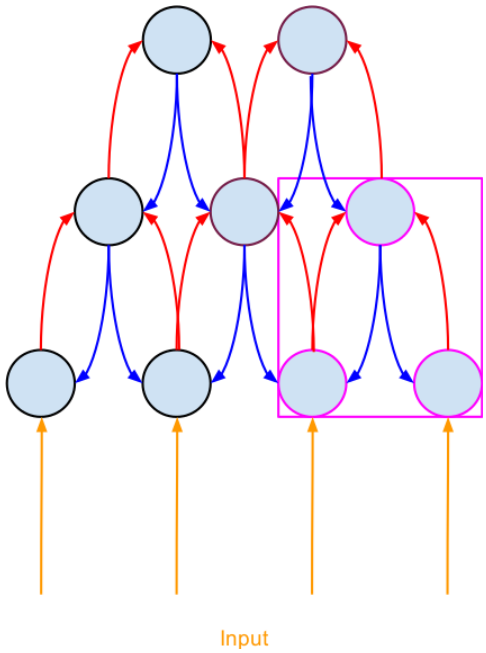
Attentional sets have not been manipulated in previous work on perceptual learning in speech perception, but some work has been done on how individual differences in attention control can impact perceptual learning. Scharenborg et al. (2014) presents a perceptual learning study of older Dutch listeners in the model of Norris et al. (2003). In addition to the exposure and test phases, these older listeners completed tests for hearing loss, selective attention, and attention-switching control. They 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 initial 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), but the fact that attention-switching control in older adults was a significant predictor of the size of perceptual learning effects (Scharenborg et al., 2014) suggests that listeners are indeed switching between sets through an experiment. The lexical decision task is oriented toward 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 adopted this attentional set for more of the exposure 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 switch attention to a more perception-oriented set could have allowed those listeners to prevent over-generalization, leading to a smaller perceptual learning effect for participants with better attention-switching control.

As stated above, Bayesian models account well for 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

Figure 1.2: 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.



learning. The predictive coding framework (Clark, 2013) provides a gain-based attentional mechanism. 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 performance where spatial resolution is poor, but attention to central

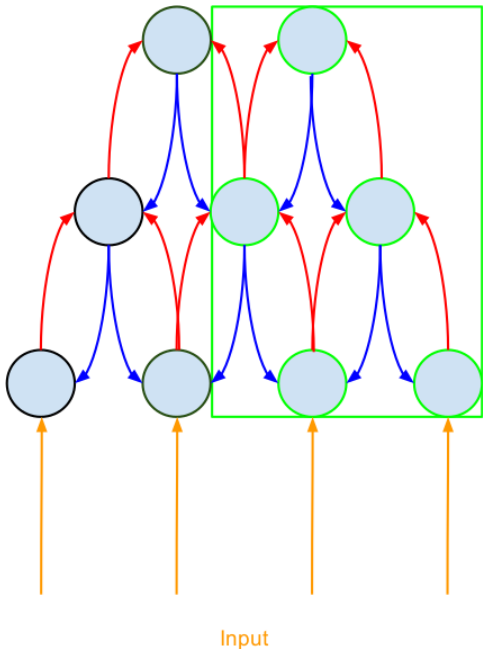
locations, where spatial resolution is high, actually harms performance (Yeshurun and Carrasco, 1998). This detrimental effect is an instance of missing the forest for the trees, as spatial resolution increased too much in the central locations to perceive the larger texture. The attentional mechanism 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 is 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. Schema for predictive coding expectation updating under a perception-oriented attentional set and under a comprehension-oriented attentional set are shown in Figures 1.2 and 1.3, respectively. In contrast, according to the mechanism proposed in Clark (2013), any increases in attention, perception-oriented or otherwise, is 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 only found 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 exposure (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.

Variability is a fundamental property of the speech signal, so sound categories

Figure 1.3: 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.



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 /str/ cluster in English words. Participants exposed to the ambiguous sound intervocalically showed a perceptual learning effect, while those exposed to the sibilants in /str/ environments did not. The sibilant in /str/ often surfaces closer to

[ʃ] in many varieties of English, due to coarticulatory effects from the other consonants in the cluster. 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 /str/ clusters, but is atypical in intervocalic position. Interestingly, given the lack of learning present in in the /str/ context, some degree of salience seems to be required to trigger perceptual learning.

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, for instance, starting as more French-like and becoming more English-like. The presentation order that showd 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 non-native speaker pronunciation changes, starting as more French-like and ending as more English-like, did not produce significantly different behaviour than control participants who only completed the categorization task. The random presentation order had perceptual learning effects in between the two ordered condtions. 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 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 had 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, and input too ambiguous given learned distributions is discarded.

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. Simply, 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.

1.5 Current contribution

Perceptual learning in speech perception 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). The two paradigms promote different attentional sets, with speech perception paradigms providing a focus on comprehension and psychophysics tasks giving focus to perception. Indeed, visually-guided perceptual learning in speech perception, with its emphasis on perception, shows largely similar exposure-specificity effects as the psychophysics findings (Reinisch et al., 2014). This dissertation expands 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, hypothesized to facilitate perceptual learning and generalization. Two experimental manipulations promote the use 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). 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). Both of these manipulations will be present in Experiments 1 and 2.

The two experiments differ in how typical 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 farther 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 predicts 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 as 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 perceptual 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 experiments, 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 (*cement*), with no /ʃ/ neighbour (*shement*), 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/ neighbour (*tashel*), 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).

Given the difference in word response rates based on position in the word (Pitt and Szostak, 2012), we predict that listeners exposed to ambiguous sounds earlier in words will be less likely to accept these productions as words as compared to listeners exposed to ambiguous sounds later in words. In addition, given the reliance of perceptual learning on lexical scaffolding, this lower acceptance rate for the Word-initial group should lead to a smaller perceptual learning effect as compared to the Word-medial group.

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. The data from 77 non-native speakers of English and two native speakers of English with reported speech or hearing disorders were excluded from the analyses. This left data from

Table 2.1: 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)

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

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. Half of the critical items had an /s/ in the onset of the first syllable (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 2.1

Four monosyllabic minimal pairs of voiceless sibilants were selected as test items for categorization (*sack-shack*, *sigh-shy*, *sin-shin*, and *sock-shock*). Two of the pairs had a higher log frequency per million words (LFPM) from SUBTLEXus

Table 2.2: 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

(Brysbaert and New, 2009) for the /s/ word and two had higher LFPM for the /ʃ/ word, as shown in Table 2.2.

All words and nonwords were recorded by a male Vancouver English speaker in 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

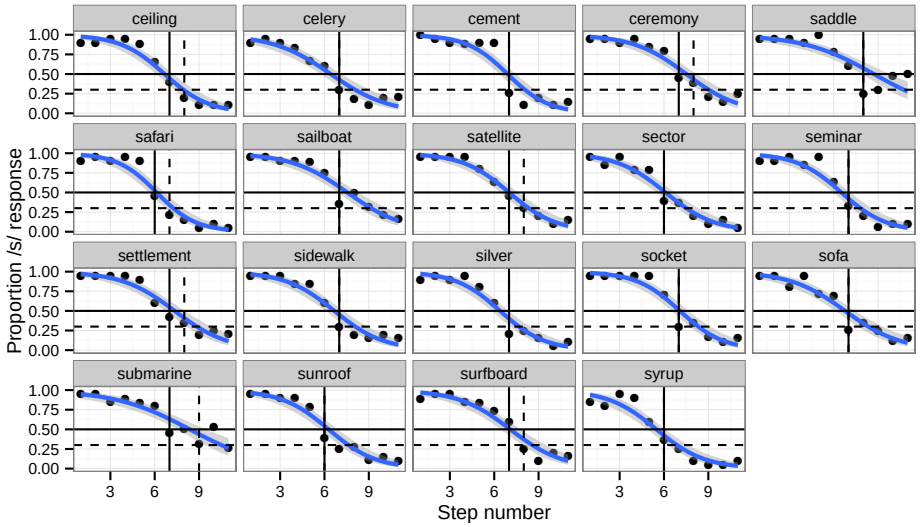


Figure 2.1: Proportion of word-responses for Word-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line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

pair continuum, resulting in 495 trials (40 exposure words plus five minimal pairs by 11 steps). The experiment was implemented in E-prime (Psychology Software Tools, 2012).

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 is in Table 2.3 and Table 2.4, 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

Table 2.3: 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

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 confirm the desired effects, a multidimensional scaled plot of acoustic distance was constructed. Using the `python-acoustic-similarity` package (Mcauliffe, 2015), sibilants were transformed into arrays of mel-frequency cepstrum coefficients (MFCC), and then pairwise distances were computed via dynamic time

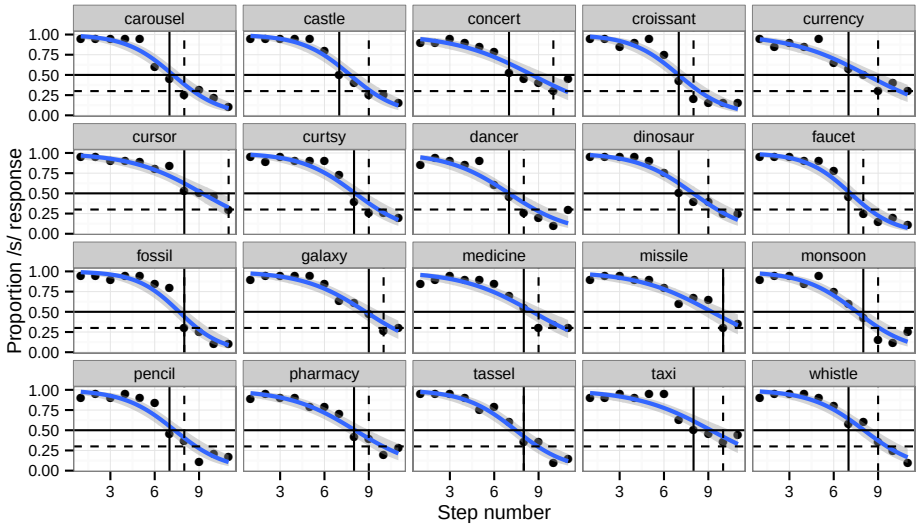


Figure 2.2: Proportion of word-responses for Word-medial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line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

warping to create a distance matrix of the sibilant productions. This distance matrix 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 principal 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 over-

Table 2.4: Step chosen for each Word-medial stimulus in Experiment 1 and the proporation /s/ response in the pretest

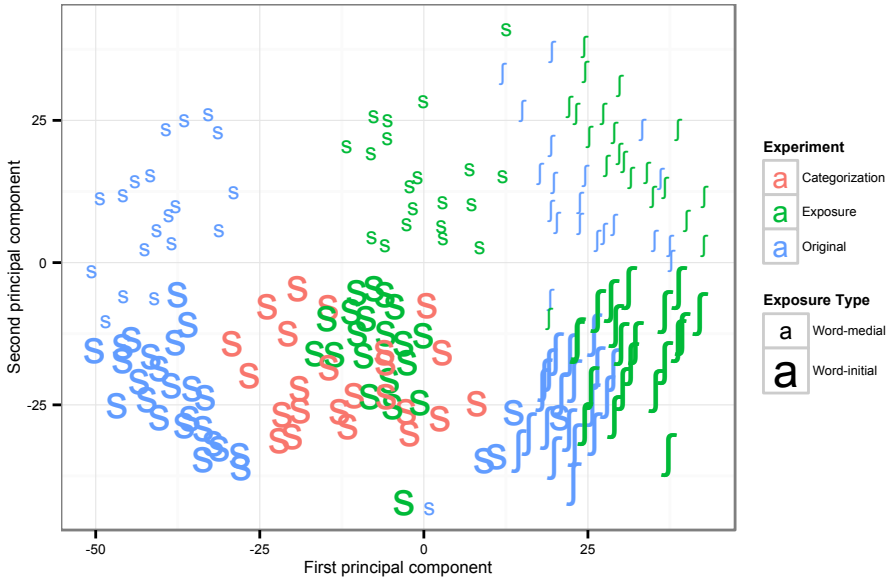
Word	Step chosen	Proportion /s/ response
carousel	7	0.45
castle	7	0.50
concert	7	0.53
croissant	7	0.42
currency	7	0.58
cursor	8	0.53
curtsy	8	0.40
dancer	7	0.45
dinosaur	7	0.50
faucet	7	0.45
fossil	8	0.30
galaxy	9	0.47
medicine	8	0.55
missile	10	0.30
monsoon	8	0.42
pencil	7	0.45
pharmacy	8	0.42
tassel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

lapping 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 2.3 and 2.4.

Procedure

Participants in the experimental conditions completed an exposure task and a categorization task. The exposure task was a lexical decision task, where participants

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).



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, with no critical or control items appearing in the first six trials and no critical or control trials in a row (following Reinisch et al., 2013).

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. Participants were instructed that there would two tasks in the experiment, and both tasks were explained at the beginning to remove experimenter interaction between exposure and categorization.

Participants were assigned to one of four conditions, plus the control experiment. Two of the conditions exposed participants to only critical items that began with /s/ (Word-initial condition) and the other two exposed them to only critical items that had an /s/ in the onset of the final syllable (Word-medial condition). This gave a consistent 200 trials in all exposure phases with identical control and filler items for all participants. Participants in half the conditions 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 assigned to the control experiment completed only the categorization task.

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 models 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 ($\beta = 0.43, SE = 0.29, z = 1.5, p = 0.13$), and Step was significant ($\beta = -2.61, SE = 0.28, z = -9.1, p < 0.01$).

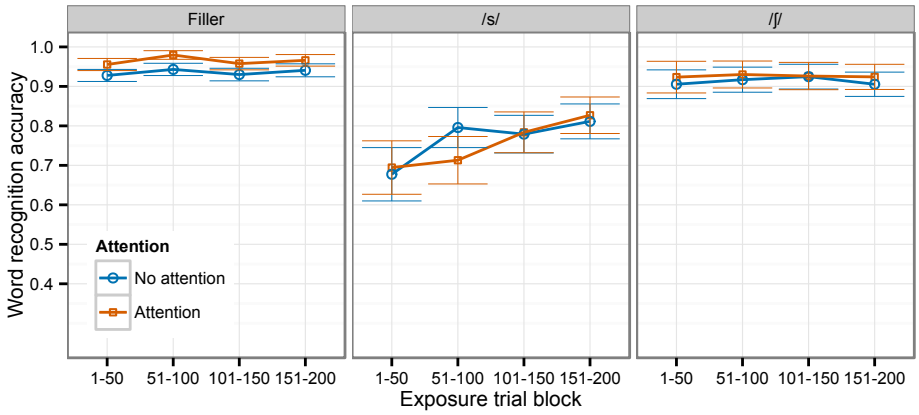
Exposure

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 high overall, with accuracy on filler trials averaging 92%. 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, and by-Subject random slopes for Trial, Trial Type, and their interactions and by-Word random slopes 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. 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 con-

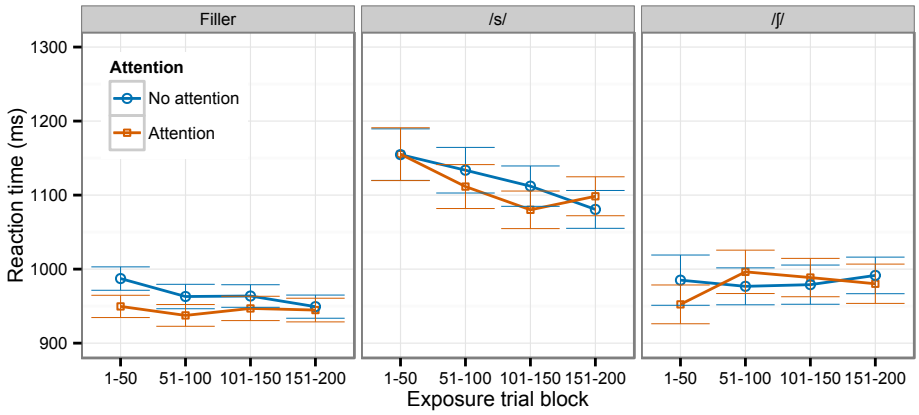
taining the modified /s/ category. Figure 2.8 shows within-subject mean accuracy across exposure, with Trial in four blocks.

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 2.9 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.

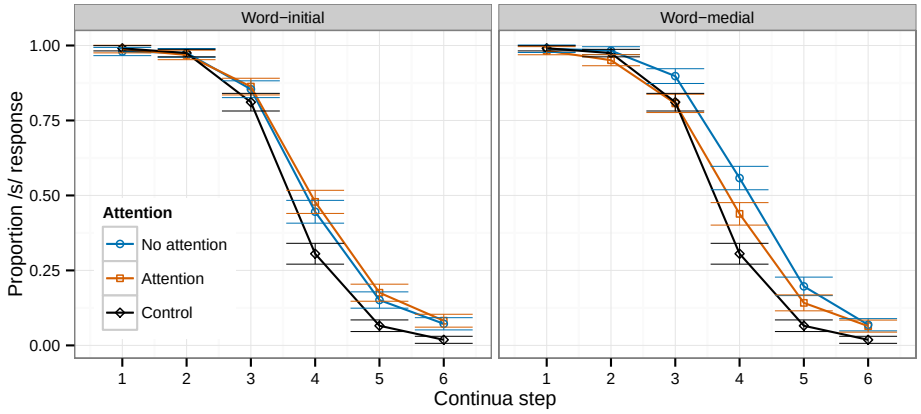


Categorization

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 crossover 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.

There was a significant effect for the intercept ($\beta = 0.76, SE = 0.22, z = 3.3, p <$

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.



0.01), indicating that participants categorized more of the continua as /s/ in general. 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 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 behaviour induced by the attention manipulation. However, when the exposure is to ambiguous /s/ tokens later in the words, we can see differences in behaviour beyond the general /s/ category expansion. Participants 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, 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 valency 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. If the valency of the attention were more positive, such as giving an explanation for the cause of ambiguity, then explicit instructions might cause a different pattern of effect. For a gain-based mechanism, however, valency of attention is not predicted to affect attention, but rather attention always increasing the gain of error signals. If valency of the instructions does

change behaviour, 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 /s/ in them are responded to slower 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 accurately 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, for instance splicing the maximally ambiguous step of the categorization continuum into expsoure 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. Exposure-specificity did not play a particularly large role in this experiment, as any effect of specificity was overriden by the experimental manipulations.

Reinisch et al. (2013) used a similar method for exposure stimuli selection

as this experiment, but generated a more typical category by using 70% word response rate as the cutoff. With these 70% stimuli, they 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 more a 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 non-native 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 2.5 and 2.6, 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.

Procedure

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

Table 2.5: 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

2.3.2 Results

Exposure

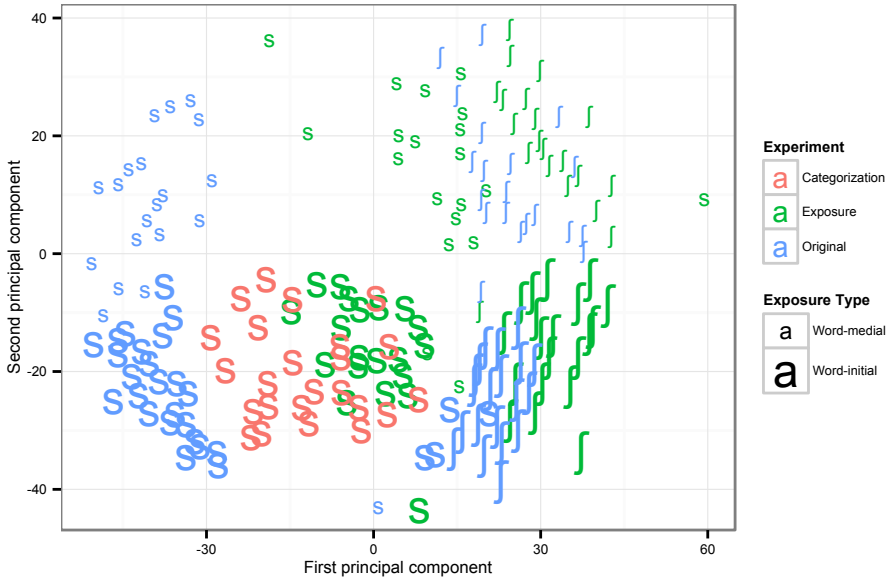
Trials with nonword stimuli and responses faster than 200 ms or slower than 2500 ms were excluded from analysis. Performance on the exposure task was as high, with accuracy on filler trials averaging 92%. A logistic mixed-effects model with accuracy as the dependent variable and a linear mixed-effects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

Table 2.6: 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

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

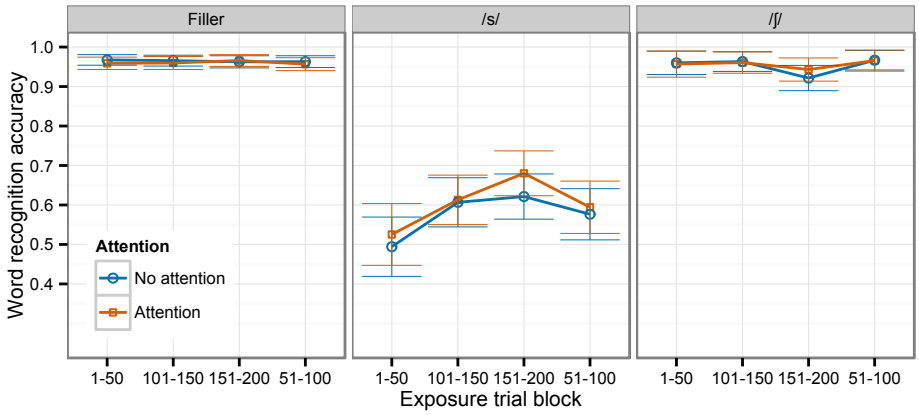
Figure 2.7: Multidimensional scaling of the acoustic distances between the sibilants of original productions, categorization tokens and the exposure tokens in Experiment 2. Categorization and exposure tokens were synthesized from the original productions using STRAIGHT (Kawahara et al., 2008).



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 ($\beta = 0.94, SE = 0.07, t = 14.4$), indicating, as in Experiment 1, 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 ($\beta = 0.07, SE = 0.02, t = 3.4$), but not the interaction between Trial and Trial Type of /s/ versus Filler ($\beta = -0.02, SE = 0.02, t = -0.8$), indi-

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.

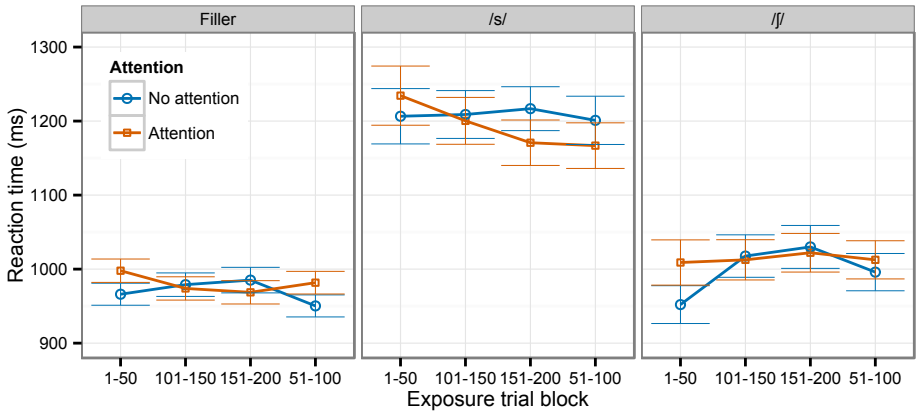


cating that reaction time remained relatively stable for words containing the modified /s/ category, but lengthened for words containing the /ʃ/ control. There was a marginal effect for Trial Type of /ʃ/ versus Filler ($\beta = 0.14, SE = 0.07, t = 1.9$), indicating that words with /ʃ/ tended to be responded to more slowly than filler times, and 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 estimated cross over point for the continuum lay outside of the 6 steps presented. A

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.



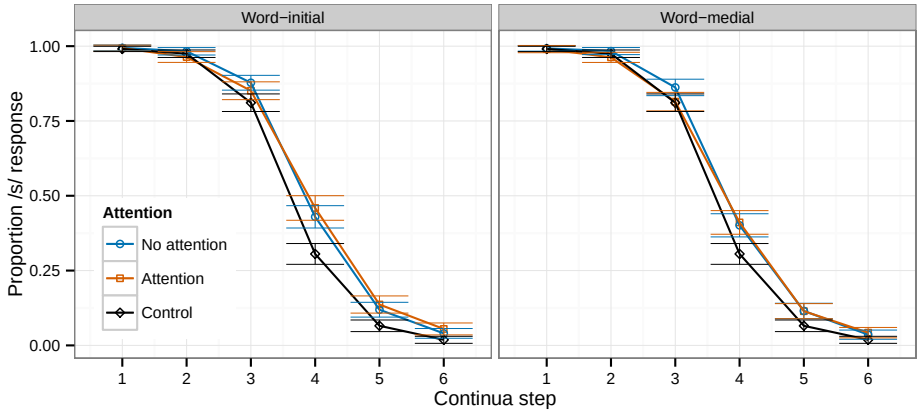
logistic mixed effects model was constructed with identical specification as Experiment 1.

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

To see what degree the stimuli used had an effect on perceptual learning, 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. In

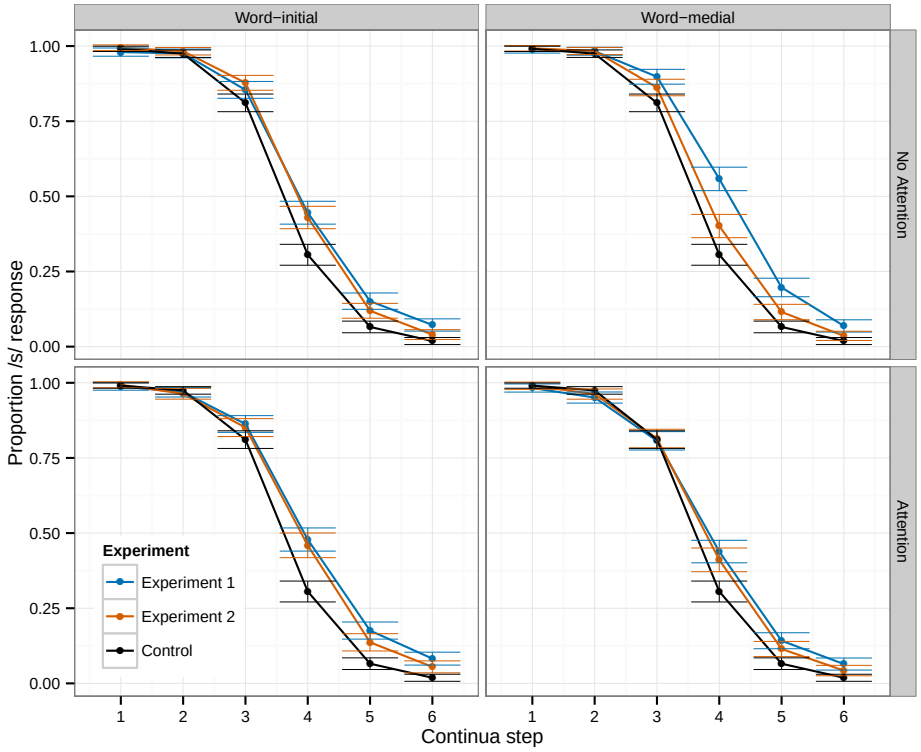
Figure 2.10: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 2. Error bars represent 95% confidence intervals.



the logistic mixed effects model, there was significant main effects for Intercept ($\beta = 1.00, SE = 0.36, z = 2.7, p < 0.01$) and Step ($\beta = -2.64, SE = 0.21, z = -12.1, p < 0.01$), 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 significant 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 potentially 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 en-

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.



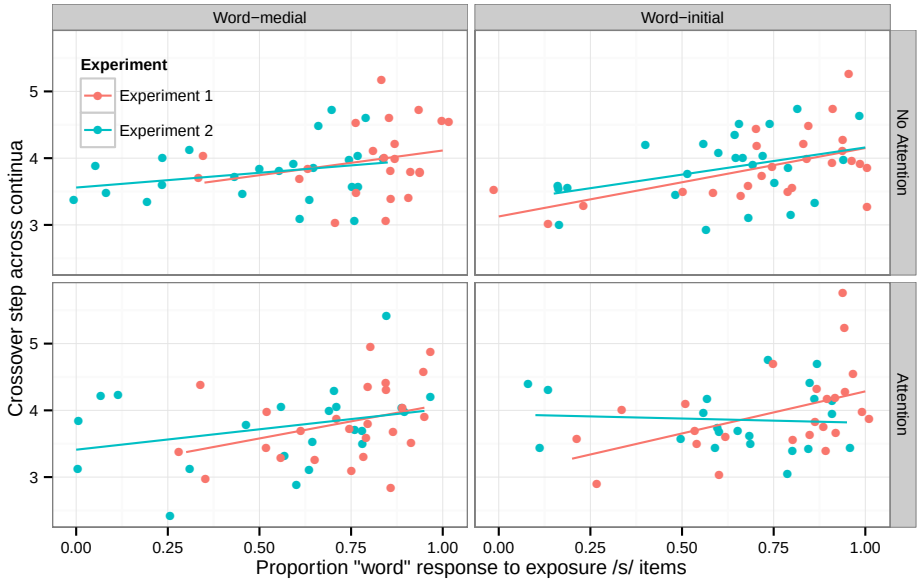
dorsement for exposure tokens and the size of perceptual learning effects (Scharenborg and Janse, 2013). Listeners who endorsed more of exposure tokens as words showed a larger perceptual learning effect. To assess such a link in the current experiments a logistic mixed-effects model was constructed identically as above, but with participants' word endorsement rate of target /s/ words as an additional fixed effect, along with its interactions with all other fixed effects. 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 endorsement 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 interactions 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 crossover points. Crossover points are where a participant's perception switches from predominantly /s/ to predominantly /ʃ/. The crossover 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 crossover 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 crossover point from /s/ to /ʃ/ in the categorization phase. Participants in the 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.

In Experiment 1, the strongest correlations between word endorsement rate and crossover 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-final/Attention condition ($r = 0.39, t(23) = 2.0, p = 0.06$). The condition for which the most perceptual learning was observed actually has the weakest correlation ($r = 0.32, t(21) = 1.5, p = 0.13$).

In Experiment 2, the strongest correlation is in the Word-initial/Attention condition ($r = 0.40, t(23) = 2.1, p = 0.05$), with two trending correlations for the

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



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 appear to fall into two categories that align with either comprehension-oriented or perception-oriented attentional sets. For most of exposure conditions, participants showed a small perceptual learning effect of similar magnitude. In these conditions, a perception-oriented attentional set was likely to be recruited. Participants who had attention

was explicitly drawn to the ambiguous sounds (Attention) showed a small perceptual learning effect Participants who were exposed to the ambiguous sounds were at the beginnings of words (Word-initial) or to the most atypical /s/ category (Experiment 2) showed similar perceptual learning effects. These manipulations were not additive. The participants exposed to the ambiguous sounds with the most lexical bias and with no attention drawn to them showed a stronger perceptual learning effect. This condition is the one where comprehension-oriented attentional sets were predicted to dominate. The perception-oriented attentional sets exhibit less generalization, similar to what is seen in psychophysics literature and in visually-guided perceptual learning in speech perception (Reinisch et al., 2014).

Compared to Experiment 1, Experiment 2 had a weaker correlation between critical word endorsement rates and crossover 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 its distance. The decreased correlation between word response rate and crossover point in the condition that saw the largest perceptual learning effects would fall out from the proposed attention mechanism. Comprehension-oriented attentional sets are proposed to update higher, more abstract sensory levels. 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 crossover point. Lexically-guided perceptual learning is induced with relatively few tokens in the literature, usually 20 of 200 total tokens (Norris et al., 2003; **?**),

but as few as modified 10 tokens have been shown to cause perceptual learning (Kraljic et al., 2008b). 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 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, as noted in Block and Siegel (2013). The propagation-limiting attention mechanism proposed earlier 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 sensory representation before updating expectations.

In and of itself, a lexical decision task biases a participant towards a comprehension-oriented attentional set. The experimental manipulations induced 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 from semantic predictability.

Chapter 3

Cross-modal word identification

3.1 Motivation

In Experiments 1 and 2, the size of the perceptual learning effects align with the elicited attentional sets. The largest perceptual learning effect was found in No Attention/Word-medial condition, which is the condition that was the least likely to promote a perception-oriented attentional set in listeners. As the lexical decision task is oriented towards comprehension, those in the No Attention/Word-medial group would have maintained a comprehension-oriented 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 words containing a modified /s/ category, semantic predictability is used in conjunction with lexical bias to boost linguistic expectations. Words that are highly predictable from context generate greater linguistic biases for the word

Semantic predictability refers to how predictable the final word in a sentence is

(Kalikow et al., 1977). 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 reduction independent of lexical factors like frequency and neighbourhood density (Scarborough, 2010; Clopper and Pierrehumbert, 2008). In speech perception work, semantic predictability and lexical bias 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). In phoneme restoration tasks, semantic predictability increases the bias to respond that a word is intact and listeners also show sensitivity to noise addition versus noise replacement for semantically predictable words (Samuel, 1981).

Samuel (1981) proposes that high predictability sentences are lower cognitive load (i.e. easier to process). The lower cognitive load allows for more attentional 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 perception-oriented. Lower cognitive load may free attentional resources which are then allocated to the task at hand. In a comprehension-oriented task, lower cognitive load would not necessarily always result in greater perceptual sensitivity. If a listener's goal is not perception, then there is no motivation for them to direct their attention to perception.

There are many possible outcomes for perceptual learning in this experiment.

Because it concerns sentence processing, many theoretical frameworks do not make explicit predictions about how the perceptual system will be updated. All theories and experimental evidence predict that ambiguous productions in predictable contexts should be endorsed more readily. But what happens to the perceptual system after that endorsement? Are the auditory tokens encoded as faithfully for sentences as for words in isolation? Even if these tokens are encoded, are they reliable enough evidence for perceptual learning? The experiment reported here takes a first pass at answer some of these questions and sets the stage for future inquiry.

One promising avenue for exploring the research questions of this chapter 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 linguistic context that conditions greater variability, such as for /s/ in /str/ clusters, then modified tokens in those contexts will not cause perceptual learning (Kraljic et al., 2008a). 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. Extending that argument 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, but not 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 attention the speaker would have directed to it. WHY A listener would view tokens that were produced more clearly, or with great care, as more reliable tokens for that speaker. Extending the argument by Kraljic and colleagues 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 and show less perceptual learning.

This all not to say that sentences will always be less effective in driving perceptual learning than words in isolation. From the literature on perceptual learning of foreign accents, sentences are extremely useful in learning to perceive 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. Unlearning effects in Kraljic and Samuel (2005) suggest that this might be asymmetric. That is, listeners may be biased towards their initial expectations for a category of native speakers in ways that they are not for nonnative speakers.

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 one paper cited in Clark (2013), participants who were expecting to see a face in static 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 (Egner et al., 2010). Without any reliability weighting or attribution of error signals, the predictive coding framework would predict larger perceptual learning effects for participants exposed to the modified category in higher predictability sentences.

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 non-native 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. A full list of words and their contexts can be found in the appendix. 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.

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 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 sentence was morphed. The preceding words in the sentence were kept as the natural production to minimize artifacts of the morphing 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.

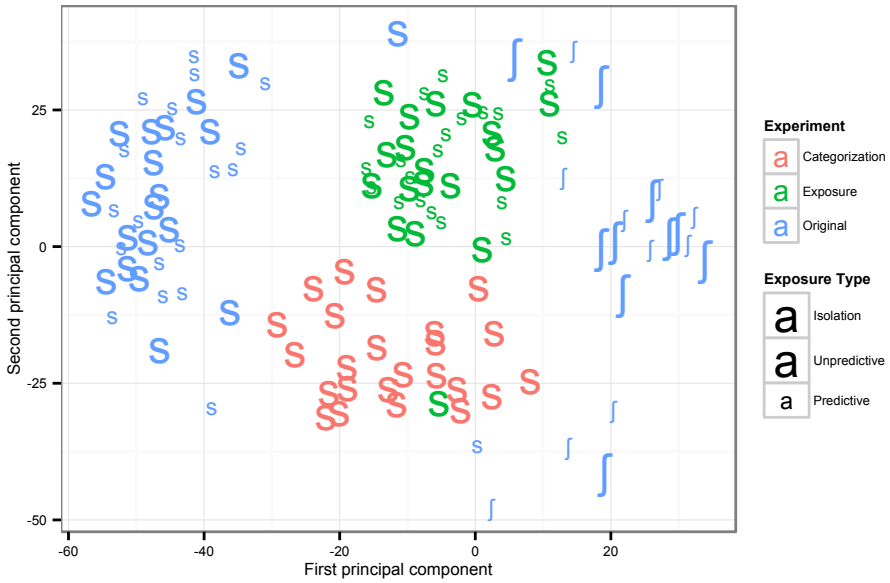
Acoustic distances between exposure tokens, categorization tokens, and their original productions were multidimensionally scaled. In Figure 3.1, the original productions are separated again by the first dimension, which corresponds to the centroid frequency of the sibilant. The categorization tokens are predictably in between the original productions, and offset in the second dimension, due to their different position in the word. The exposure tokens for Experiment 3 fit in between the original productions and the categorization tokens.

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 behaviour 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.

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

Figure 3.1: Multidimensional scaling of the acoustic distances between the sibilants of original productions, categorization tokens and the exposure tokens in Experiment 3.



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 sentence fragments that were lacking in the final word. They 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.

The measure used for determining rewriting of sentences was the proportion of participants that included the target word in their responses. For predictive sentences, the mean proportion was 0.49 (range 0-0.95) and for unpredictive sentences, the mean proportion was 0.03 (range 0-0.45). Predictive sentences that had target response proportions of 20% 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 ambiguities.

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 Procedure

As in Experiments 1 and 2, participants completed an exposure task and a categorization task. For the exposure task, participants heard a sentence via headphones for each trial. Immediately following the auditory presentation, they were pre-

sented 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.

Participants were assigned to one of four groups of 25 participants. In the exposure phase, half of the participants were exposed to a modified /s/ sound only in Predictive sentences and half were exposed to it only in Unpredictive sentences. Half of all participants were told that the speaker's production of "s" was sometimes ambiguous, and to listen carefully to ensure correct responses.

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

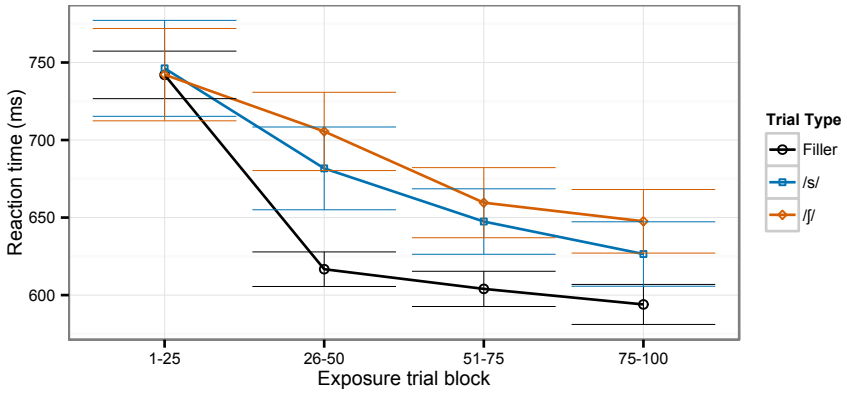
3.3 Results

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.



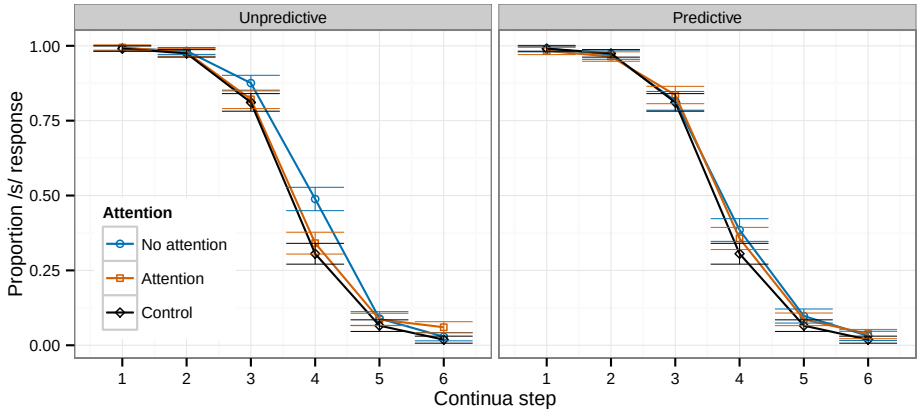
A significant effect was found for Trial ($\beta = -0.20, SE = 0.01, t = -11.0$), indicating that reaction time became faster over the course of the experiment. There was a significant effect for Trial Type of /ʃ/ versus Filler ($\beta = 0.19, SE = 0.09, t = 2.1$), but not for /s/ versus Filler ($\beta = 0.11, SE = 0.09, t = 1.3$), 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 reactions times.

Participants were faster in this task than in the lexical decision task.

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).

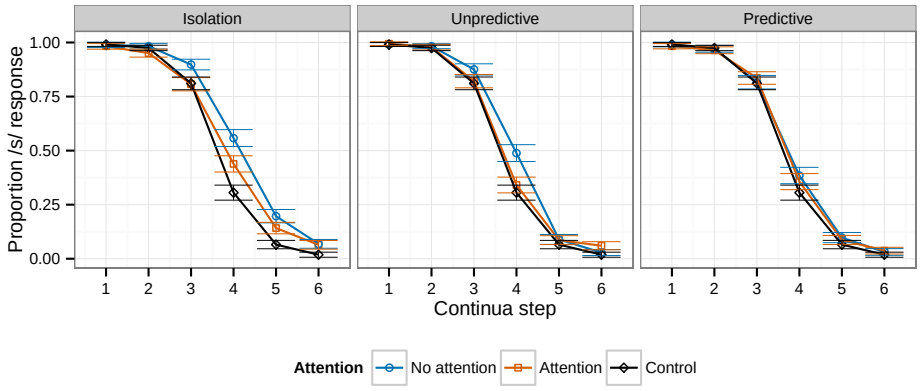
Figure 3.3: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 3. Error bars represent 95% confidence intervals.



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 ($\beta = 0.23, SE = 0.23, z = 0.97, p = 0.33$), Atten-

tion ($\beta = 0.30, SE = 0.21, z = 1.4, p = 0.15$), 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). The means for the center steps in the Unpredictive condition appear to be different depending on Attention, while those in the Predictive condition show no such Attention difference. As will be shown later, the lack of an effect is likely due a lack of statistical power, suggesting that the difference between the Exposure Type conditions is a smaller than the crucial effect 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 set data 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 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 sharper crossover 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 percep-

tual learning effect, and participants in the Unpredictive condition were in between Predictive and Isolation, but not significantly different from either. Increasing the statistical power would be necessary to separate the conditions out 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. Particularly, as the predictability of the context increases, the amount of perceptual learning decreases. 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 is 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 appears to be a linguistic source of variation and shows identical effects as a more local source like consonant cluster coarticulation.

The prediction of a simple predictive coding model (Clark, 2013) were 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.

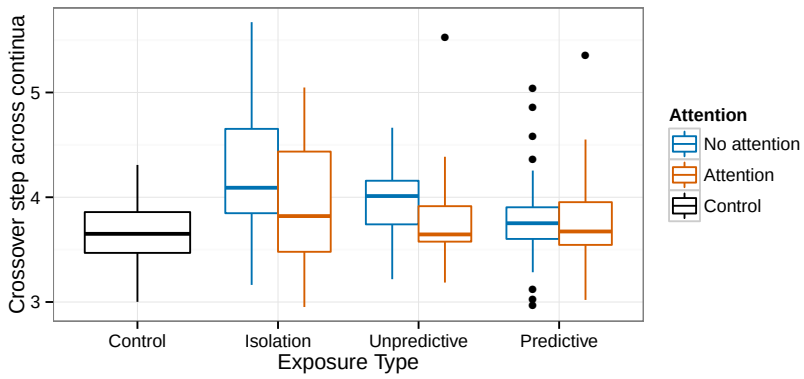
However, if semantic predictability functions in an identical way as consonant cluster coarticulation, listeners would not perceptual learning effects even if they were tested on a continuum in a high predictability sentence. In Kraljic et al. (2008a), listeners exposed to an ambiguous /s/ in the context of /str/ and then tested on a continuum from /astri/ to /aʃtri/ showed no perceptual learning effects. Participants who were exposed to ambiguous /s/ intervocalically showed perceptual learning on both /asi/-/aʃi/ and /astri/-/aʃtri/ continua. In this case, there is no exposure-specificity effects, so participants do 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.

One question raised by this finding is whether any and all variation can be attributed to these contextual factors. For instance, if this experiment had used the ambiguity threshold of Experiment 2 instead of Experiment 1, would a different pattern of results emerge? If high predictability only resulted in broadening of perceptual categories, participants in the Predictive condition should show nearly as much perceptual learning as the current participants in the Unpredictive condition. Increasing the atypicality of the category would likely lead to reduced perceptual learning for future participants in the Unpredictive category, mirroring the effects

between Experiments 1 and 2.

As a final point in this discussion, the distribution of individuals' perceptual learning effects differs across conditions. Figure 3.5 shows the distribution of crossover points of each subject in Experiment 3 and participants in the condition of Experiment 1 that use the same exposure words. Crossover points are where along the continua the participant switches their perception from primarily /s/ to primarily /ʃ/. Participants in the Predictive/No Attention condition have the most unique distributional pattern. In this condition, the distribution of crossover points is fairly constrained, with a narrow interquartile range, but with many outliers. This distribution is more peaky with long skirts to either side. For other conditions, the distributions are captured well by the boxplots, with relatively large interquartile ranges. These distributions are more broad, but narrower skirts. Participants in the Predictive/No Attention could potentially be more separable into discrete groups than the other conditions where it may be more continuous.

Figure 3.5: Distribution of crossover points for each participant across comparable exposure tokens in Experiments 1 and 3.



One possible reason for these more discrete groups could have to do with 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 attentional resources 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, revealing a relationship as was found in Scharenborg and Janse (2013).

Chapter 4

Discussion and conclusions

This dissertation set out to examine the influence of listener attentional sets on perceptual learning. Perceptual learning is phenomenon common to many fields involved in cognitive science. How perceptual learning manifests, however, is quite different. Perceptual learning in psychophysics is the process of a perceiver aligning their senses to the world. Perceptual learning in speech perception is process by which perceivers align their perceptual system to an interlocutor to facilitate understanding. That final part is crucial. In speech perception, truly accurate perception is immaterial if a listener can understand enough to interact fluently with their interlocutor.

I argue that the perceptual learning as a mechanism is shared between linguistic and non-linguistic domains. The goals of the tasks and the attentional sets promoted by them are where differences in patterns of behaviour arise. Perception-oriented tasks in all domains leads to less generalized learning. Comprehension-oriented tasks lead to more generalized learning. The first two experiments of this dissertation implement a standard lexically-guided perceptual learning paradigm,

but with manipulations to promote perception-oriented attentional sets. Even in a comprehension-oriented 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 relate to specificity and generalization in the perceptual learning literature. The four manipulation to promote attentional sets are then examined, followed finally by implications for models of the brain 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 when they were exposed to ambiguous sounds later in the words rather than at the beginning of words. 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-specificity at the level of sound in context (i.e. word-initial /s/) was tested, but no such specificity was found. Exposure to /s/ in word-medial position resulted in perceptual learning effects at least as large as for exposure to word-initial /s/. Perceptual learning 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. The results of this dissertation show that not only do listeners show perceptual learning effects

from stimuli with large degrees of coarticulation (i.e., in the middle of the word) to stimuli without as much coarticulation, in some cases, perceptual learning is greatest in precisely the cases where acoustic distance is greatest. 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. While visually-guided perceptual learning shows comparable effects to lexically-guided perceptual learning (van Linden and Vroomen, 2007), 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). This dissertation argues for a mechanism of attention that limits updating of expectations to the level at which attention is directed. If attention is oriented towards perception, then expectations must be resolved at a low level of sensory representation. At this level, fine details of the sensory input are encoded, and generalization to other stimuli

is more difficult due to the fine-grained encoding. If attention is oriented toward comprehension, expectations can be resolved at a higher level with more abstract sensory representations. At this level, fine details are not encoded, allowing generalization to other stimuli more readily. A consequence of this mechanism nicely captures the differences in numbers of stimuli 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 sensory representation will generalize more than many repetitions updating fine-grained sensory representations. From this, we could predict that word endorsement rate and category boundary shift would be the less (linearly) correlated the more comprehension-oriented participants are. This prediction is borne out by the lack of correlation between word endorsement rate and crossover 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 strong correlation between word endorsement rate and crossover point is not present. Participants in this condition have relatively high crossover 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 per-

ceptual learning is due to increased maintainance 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 these 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 changed. 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 behavioural 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. Semantic predictability has previously been shown to affect perception-oriented tasks a similar way that lexical bias affects them(Connine, 1987; Borsky et al., 1998) In Experiment 3, 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 /str/ cluster that conditions that variation (Kraljic et al., 2008a). 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 /str/ clusters, with the /s/ produced more in a postalveolar position due to the upcoming /r/. 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 there would be a simple relaxing of all auditory categories.

Would semantic predictability have the same effect on nonnative speakers? Several studies have shown perceptual learning benefits for multiple talkers of an accent using sentential stimuli (Bradlow and Bent, 2008, and others). Clearly, perceptual learning of accents is possible through hearing sentences of varying predictability, but the phonetic variability involved in those tasks reaches far beyond that involved here. The speaker producing the sentences in this dissertation is a native English speaker. 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.

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, 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 to these stereotype linguistic variables would be inhibited relative to other variables. 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 progres-

sively 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). Would less salient social variants be learned easier?

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 crossover 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.

4.5 Implications for cognitive models

The primary model that this dissertation adopts is that of the predictive coding framework (Clark, 2013). In this model, expectations about sensory input are fed from higher levels of representation to lower ones. The mismatch between actual sensory input 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 well 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, schematized in Figures 1.2 and 1.3. Such a mechanism explains both the previous findings and the current results.

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. Increasing linguistic expectations through increased lexical bias induced more comprehension-oriented attentional sets, resulting in larger perceptual learning effects. Increasing semantic predictability resulted in no perceptual learning effects, likely due to the attribution of the modified sound category to reduced speech clarity. Inducing perception-oriented attentional sets through explicit instructions or making the modified sound category more atypical resulted in small, but robust perceptual learning effects. 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. Given the robust perceptual learning effects across all conditions, perceptual learning is a largely automatic process.

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 3

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, 10

Block, N. and Siegel, S. (2013). Attention and perceptual adaptation. *The Behavioral and brain sciences*, 36(3):205–6. → pages 23, 58, 86

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 18, 61, 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 18, 61

Bradlow, A. R. and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2):707–729. → pages 2, 63, 83

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 33

Clare, E. (2014). Applying phonological knowledge to phonetic accommodation. → pages 17

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and brain sciences*, 36(3):181–204. → pages x, 6, 7, 10, 11, 12, 23, 24, 44, 58, 64, 74, 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 17, 61

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 18, 61, 82

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1987). Phoneme identification and the lexicon. → pages 3, 14, 20, 80

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 63

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 : the official journal of the Society for Neuroscience*, 30(49):16601–16608. → pages 64

Eimas, P. D., Cooper, W. E., and Corbit, J. D. (1973). Some properties of linguistic feature detectors. → pages 8

Eisner, F. and McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2):224–238. → pages 2, 11, 45

Eisner, F. and McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4):1950. → pages 10

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 18, 61

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110–125. → pages 14

Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological bulletin*, 50(6):401–431. → pages 2, 8, 19

Gilbert, C., Sigman, M., and Crist, R. (2001). The neural basis of perceptual learning. *Neuron*, 31:681–697. → pages 11, 24, 28

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 63

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* . . . , 61(5). → pages 17, 61, 65

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 *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3933–3936. → pages xii, 33, 38, 50, 65

Kleber, F., Harrington, J., and Reubold, U. (2012). The Relationship between the Perception and Production of Coarticulation during a Sound Change in Progress. → pages 55

Kleinschmidt, D. and Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. *Science*, (June):10–19. → pages 6, 10, 11, 86

Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008a). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, 107(1):54–81. → pages 17, 25, 62, 74, 75, 80, 83

Kraljic, T. and Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, 51(2):141–78. → pages 10, 11, 12, 17, 45, 64

Kraljic, T. and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1):1–15. → pages 2, 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 2, 17, 26, 58, 62, 63, 64, 74, 80

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. → pages 84

Ladefoged, P. and Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1):98–104. → pages 2

Leber, A. B. and Egeth, H. E. (2006). Attention on autopilot: Past experience and attentional set. → pages 20, 22

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177. → pages 19

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 18

Mattys, S. L. and Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160. → pages 15, 18, 21, 61, 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 : JSLHR*, 40(3):686–693. → pages 18, 61

Mcauliffe, M. (2015). python-acoustic-similarity. → pages 35

Mitterer, H., Scharenborg, O., and McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2):356–361. → pages 12, 80

Norris, D. and Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & psychophysics*, 43(6):541–550. → pages 20

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. → pages 2, 5, 6, 9, 10, 12, 14, 17, 21, 24, 28, 45, 57, 80

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*. → pages 58, 85

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, pages 137–158. → pages 27

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, 15, 17, 20, 21, 29, 30, 31

Pitt, M. A. and Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory & Language*, 29(5):611–632. → pages 20

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 16

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, 15, 29

Psychology Software Tools, I. (2012). E-Prime. → pages 34, 66, 67

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, 24, 38, 39, 45

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, 24, 28, 57, 58, 80

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of experimental psychology. General*, 110(4):474–494. → pages 16, 18, 27, 61

Samuel, A. G. (1986). Red herring detectors and speech perception: in defense of selective adaptation. *Cognitive psychology*, 18(4):452–499. → pages 8

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory phonology*, pages 575–604. → pages 17, 61, 63

Scharenborg, O. and Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, perception & psychophysics*, 75(3):525–36. → pages 22, 54, 77

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 21, 22, 58, 86

Shankweiler, D., Strange, W., and Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. *Perceiving, acting, and knowing: Toward an ecological psychology*, pages 315–345. → pages 1

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119(1):131–136. → pages 26, 27, 30, 63, 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):EL485. → pages 1

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 85

van Linden, S. and Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of experimental psychology. Human perception and performance*, 33(6):1483–1494. → pages 10, 80

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 2, 8, 9, 11, 26, 27, 58

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 20

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396(6706):72–75. → pages 24, 58