Effect of online processing on linguistic memories

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

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!

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 of a language are faced with a large degree of phonetic variability when interacting with their fellow language users. Speakers can have different sizes, different genders, and different backgrounds that make speech sound categories, at first blush, overlapping in distribution and hard to separate in acoustic domains. In addition to properties of the speaker varying, a listener's attention or goals in an interaction can vary, such as listening hard to a non-native speaker or being distracted by planning upcoming utterances or by another task entirely. Despite variability on the part of both the listener and the speaker, listeners can maintain a large degree of perceptual constancy, interpreting disparate and variable productions as belonging to a single word type or sound category. Exposure to a speaker affects a listener's perceptual system for that speaker, increasing their ability to understand that speaker or other speakers of similar background [5]. Broadly, perceptual learning or perceptual adaptation in the speech perception literature refers to the updating of distributions corresponding to a sound category for a particular speaker.

This dissertation investigates the interaction between linguistic, attentional and acoustic factors on perceptual learning. The key hypothesis being tested is that the more evidence a listener has toward a specific word, be it acoustic, lexical or semantic/sentential, the stronger the link between an ambiguous sound embedded in that word and a sound category will be. The more strong evidence that a listener accrues for a given sound category of a speaker, the more perceptual learning will take place. This chapter will review the recent literature on perceptual learning in speech perception (Section 1.1), as well as literature on linguistic (Section 1.2), attentional (Section 1.3), and signal (Section 1.4) factors. Chapter 2 will detail two experiments using a lexically-guided perceptual learning paradigm, each with manipulations to lexical bias and attention. The two experiments differ in the acoustic properties of the exposure tokens. Chapter 3 details an experiment using a novel perceptual learning paradigm that manipulates an additional linguistic factor, namely semantic predictability, to increase the linguistic expectations during exposure. The perceptual learning literature has generally used consistent processing conditions to elicit perceptual learning effects, and the goal of this dissertation is to examine the robustness and degree of perceptual learning across different processing conditions.

1.1 Perceptual learning

Norris et al. [29] began the recent set of investigations into lexically-guided perceptual learning in speech. Norris et al. [29] 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”, and 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 on 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 them at the end of nonwords did not. The exposure using words was further differentiated by the bias introduced by the words. 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/ dependent on the bias of the word, which led to an expanded category for that fricative at the expense of the other category.

In addition to lexically-guided perceptual learning, unambiguous visual cues to sound identity can cause perceptual learning as well (referred to as perceptual recalibration in that literature). In Bertelson et al. [3], an auditory continuum from /aba/ to /ada/ was synthesized and paired with a video of a speaker producing /aba/ and a video of /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 with the video for /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 video of /aba/ paired with the ambiguous token in the preceding block, and likewise for /ada/.

van Linden and Vroomen [41] compared the perceptual recalibration effects from the visual lipread paradigm [3] to the standard lexically-guided perceptual learning paradigm [29]. Lipread recalibration and perceptual learning effects had comparable size, 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 effects did not last through prolonged testing in this study, unlike in other studies [14, 21]. In those studies, lexically-guided perceptual learning effects persisted through intermediate tasks, as long as the task did not involve any contradictory evidence of the trait learned [21], and also persisted across 12 hours [14].

Perceptual learning is a well established phenomenon in the psychology and psychophysics literature. Training can improve a participants ability to discriminate in many disparate modalities, such as visual acuity, somatosensory spatial resolution, weight estimation, and discrimination of hue and acoustic pitch [16, for review]. In this literature, perceptual learning is the improvement of a perceiver to judge the physical characteristics of objects in the world through training that assumes attention on the task, but doesn't 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 for a sound category, narrow their own perceptual category, resulting in a change in variance of the category, not of the mean of the category (along some acoustic-phonetic dimension) [12, 36, 43]. Perceptual learning or recalibration in the speech perception literature is a more broad updating of perceptual categories, either in mean or variance [29, 43].

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 [17]. Perceptual learning in speech perception has shown a large degree of exposure-specificity, where participants do not generalize cues across speech sounds [35] or across speakers unless the sounds are similar across exposure and testing [13, 21, 22, 33]. On the other hand, 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 generally quite different from the exposure stimuli, with participants exposed to multisyllabic words ending in an ambiguous sound and tested on monosyllabic words [34] and nonwords [21, 29], though exposure-specificity is found when exposure and testing use different positional allophones [28].

Perceptual learning can be captured well in terms of Bayesian belief updating [20] or as part of a predictive coding model of the brain [8]. In Bayesian belief updating, the model categorizes the incoming stimuli based on multimodel cues, and then updates the distribution to reflect that categorization. This updated conditional distribution is then used for future categorizations in an iterative process. Kleinschmidt and Jaeger [20] model the results of the behavioural study in Vroomen et al. [43], with models fit to each participant capturing the perceptual recalibration and selective adaptation shown by the participants over the course of the experiment. A similar, but more broad, framework is that of the predictive brain [8]. 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.

1.2 Linguistic factors and perceptual learning

The two linguistic factors manipulated in this dissertation are lexical bias and semantic predictability. Lexically-guided perceptual learning paradigms use lexical bias as the means to link an ambiguous sound to a sound category. In Chapter 3, a novel, sententially-guided perceptual learning paradigm is used to manipulate the listener's linguistic expectations in a different manner than manipulating lexical bias.

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 one sound, listeners in general are more likely to interpret any step along the continuum as the word endpoint rather than the nonword endpoint [15]. This bias is exploited in perceptual learning studies to allow for noncanonical, ambiguous productions of a sound to be linked to pre-existing sound categories. Given that ambiguous productions must be associated with a word to induce lexically-guided perceptual learning [29], differing degrees of lexical bias could lead to differing degrees of perceptual learning, a prediction which will be tested in Chapter 2. The stronger the lexical bias, the stronger the link will be between the ambiguous sound and the sound category, as mediated by the word.

Studies looking at lexical bias generally use a phoneme categorization task, where participants identify a sound at the beginning or end of a word from two possible options that vary in one dimension. For instance, given a continuum from /t/ to /d/, participants are asked to identify the sound at the beginning or end as either /t/ or /d/. Lexical bias effects are calculated based on two continua, where words are formed at opposite ends. For the /t/ to /d/ continua, one would form a word at the /t/ end, such as *task* and one would form a word at the other end, such as *dash*. Lexical bias effects are then calculated from the different categorization behaviour for these two continua.

Connine and Clifton [11] established different reaction time profiles for "perceptual" and "postperceptual" processes in categorizing a continuum. With lexical bias, response times to the end points of a /d/ to /t/ continuum show no difference whether the continuum forms a word at one end point (such as *dice* to *tice*) or at the other (such as *dype* to *type*). However, at the boundary between /d/ and /t/, reaction times are faster when a subject responds consistent to the bias (i.e. interprets an ambiguous word *?ice* as *dice*). For postperceptual processes, in this case a monetary payoff for responding either /d/ or /t/ along a continuum that has nonwords at both ends, the pattern is reversed, where reaction times were faster for responses consistent with the monetary bias at the end points of the continuum, but no such difference was found for the category boundary. Both biases produced similar categorization patterns, such that the participants biased toward /d/, either lexically or monetarily, categorized more of the continuum as /d/, so the principle difference between the two biases was in the reaction time profile.

Lexical bias varies in strength according to several factors. First, the length of the word in syllables has a large effect on lexical bias, with longer words showing stronger lexical bias than shorter words [32]. Continua formed using trisyllabic words, such as *establish* and *malpractice*, were found to show consistently large lexical bias effects than monosyllabic words, such as *kiss* and *fish*. Pitt and Samuel [32] also found that lexical bias from trisyllabic words was robust across experimental conditions, but lexical bias from monosyllabic words was more fragile and condition dependent. They argue that these affects arise from both the greater bottom-up information present in longer words and the greater lexical competition for shorter words.

Pitt and Szostak [30] used a lexical decision task with a continuum of fricatives from /s/ to /ʃ/ embedded in words differing in the position of a sibilant. They found that ambiguous fricatives earlier in the word, such as *serenade* or *chandelier*, lead to greater nonword responses than the same ambiguous fricatives embedded later in a word, such as *establish* or *embarass*. In the experiments and meta-analysis of phoneme identification results presented in Pitt and Samuel [31], they found that, for monosyllabic word frames, token-final targets produce more robust lexical bias effects than token-initial targets.

In the stimuli used in Norris et al. [29] and most other lexically-guided perceptual learning experiments, lexical bias tends to be maximized. Exposure stimuli are multisyllabic words that end in the ambiguous sound. Perceptual learning has also been found when the ambiguous stimuli is embedded earlier in the word, such as the onset of the final syllable [21, 23, 24] or the even the onset of the first syllable [7]. In light of the findings in Pitt and Szostak [30], we would expect different lexical biases at the point that those ambiguous sounds were heard, and therefore, I hypothesize, different endorsement rates and different perceptual learning effect sizes. This prediction will be explicitly tested in Experiment 1 in Chapter 2.

1.2.2 Semantic predictability

The second type of linguistic expectation manipulation used in this dissertation is known as semantic predictability [18]. Sentences are semantically predictable when they contain words prior to the final word that points almost definitively to the identity of that final word. For instance, the sentence fragment *The cow gave birth to the...* from Kalikow et al. [18] 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, other than having the category of noun.

Perceptual learning paradigms for specific native speaker characteristics restrict themselves to lexical biases, but we predict compounding effects for perceptual learning, from studies done in speech production [9, 37]. Studies looking at perceptual adaptation to nonnative accents have used exposure tasks that incorporate linguistic information beyond the lexical domain. For instance, **?**] and Bradlow and Bent [5] trained listeners on foreign-accented English using sentence exposure items, and **?**] trained listeners on a merger of /i/ and /ɪ/ in French-accented English using a multi-sentence story. In these studies, while the amount of linguistic information available is greater, so too are the number of characteristics that need to be learned, if the listener does not already have training or experience with the specific non-native accent.

In speech production, higher semantic predictability has been found to result in acoustically reduced word tokens. In Scarborough [37], words produced in highly predictable frames tend to be shorter in duration and have less dispersed vowel realizations. This semantic predictability effect did not interact with neighbourhood density, a lexical factor that proxies for the amount of lexical competition a word has. Words with many neighbours, and therefore less lexical predictability, had longer durations and more dispersed vowel realizations. For both the lexical and the semantic predictability, high predictability led to less distinct word realizations, and low predictability led to more distinct word realizations, independently of each other. In a study looking at semantic predictability across dialects, Clopper and Pierrehumbert [9] found that not all dialects realize the effects of semantic predictability the same. For the Southern dialect of American English, the results were much the same as in Scarborough [37], showing temporal and spectral reduction in high predictabilty environments. However, speakers in the Midland dialect showed no such effect, and speakers of the Northern dialect showed more extreme Northen Cities shifting in the the high predictability environment.

Despite the temporal and spectral reduction found in high predictability contexts, 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 [18], which has been found across children and adults [**?**], and across native monolingual and early bilingual listeners, but not late bilingual listeners [**?**]. 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 [26]. 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 [**?**].

Two studies looking at how similar lexical and bias and semantic predictability are, found conflicting results for their reaction time profiles on a phoneme categorization task [4, 10]. Connine [10] found evidence that semantic predictability operated similar to the "postperceptual" processes in Connine and Clifton [11]. Borsky et al. [4] attempted to replicate Connine [10] while removing potential confounds from the methodological design. In contrast to Connine [10], they used only one voicing continuum (from *goat* to *coat*), embedded the acoustic target in the middle of the sentence rather than at the end, and only presented one instance of each sentence to each participant rather than all continuum steps for each sentence to each participant. The reaction time profile in their study more closely aligned to the profile of lexical bias in Connine and Clifton [11].

The previous literature on semantic predictability has shown largely similar effects as lexical bias. In the few instances where both are manipulated, they seem to be orthogonal effects that increase a listener's linguistic expectation. However, those studies were done in production [9, 37], and perception does not necessarily mirror production. It may be the case that there is a maximum to the linguistic expectation that a listener has for a particular utterance. Committing too much to a particular expectation could lead to garden path phenomena [**?**]. The effect of semantic predictability on perceptual learning will be explicitly tested in Chapter 3.

1.3 Attentional factors and perceptual learning

Attention is a large topic of research in its own right, and this review only reviews literature that is directly relevant to perceptual learning. Attention has been found to have a role on perceptual learning in the psychophysics literature. For instance, Ahissar and Hochstein [1] 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. However, there was a small degree of local detection learned from global detection, with the singleton popping out from its field as highly salient.

Attentional sets are a widely used term in the attention literature. In the visual domain, attentional sets can refer to the strategies that the perceiver uses to perform a task. For instance, in a visual search task, colour, orientation, motion and size are the predominant strategies [44]. Two broad categories of attentional sets are generally used. Focused sets direct attention to components of the sensory input, and diffuse sets direct attention to global properties of the sensory input. In the visual search literature, a focused attentional is the feature search mode, which gives priority to a single feature, such as the colour of the target, and a diffuse attentional set is singleton detection mode, which gives priority to any salient features [2]. In the auditory streaming literature, two attentional sets have been identified as "selective listening", where the perceiver attempts to hear the components of two streams, and “comprehensive listening”, where the perceiver tries to hear all components as a single stream [42]. Finally, in a lexical decision tasks, the diffuse attentional set is where primary attention is on detecting words from nonwords, and the focused attentional set is where instructions direct participants attention to a potentially misleading sound [30]. Attentional set selection is not necessarily optimal on the part of the perceiver, and it has been shown to be biased based on experience, with the amount of training performed influencing the length of time that perceivers will continue to use non-optimal sets after the task has changed [25].

Attention has 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. [39] presents a perceptual learning study of older Dutch listeners in the model of Norris et al. [29]. In addition to the exposure and test phases, these older listeners completed three 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 perecptual learning effects, which the authors ascribed to an increased reliance on lexical information. Older listeners had previously been shown to have smaller perceptual learning effects as compared to younger listeners, but the differences were most prominent directly following exposure [38]. 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 smaller initial perceptual learning effects, but no such decay. Scharenborg and Janse [38] also found that performance in the lexical decision task significantly affected the perceptual learning in the testing phase.

Lexical bias is affected by attentional processing conditions. Pitt and Szostak [30] additionally investigated the role of attention in modulating lexical bias. When listeners were told that the speaker’s /s/ and /ʃ/ were ambiguous and to listen carefully to ensure correct responses, they were less tolerant of noncanonical productions across all positions in the word. That is, participants attending to the speaker’s sibilants were less likely to accept the modified production as a word than participants given no particular instructions about the sibilants. Given that the task listeners performed was a lexical decision task, the default attentional set for the task, termed "diffuse" by Pitt and Szostak [30], would have attention distributed across both acoustic-phonetic and lexical domains. Listeners given the instructions about the speaker's /s/ productions had a "focused" attentional set, with more weighting on the acoustic-phonetic domain than the "diffuse" attentional set. Under higher cognitive load, such as performing a more difficult concurrent task, listeners show an increased lexical bias, as a result of weaker encoding of the auditory details [27]. These results suggest that detailed encoding requires attentional resources. In Mattys and Wiget [27], the primary task was a phoneme identification task, where a "focused" attentional set would likely be the default for participants, but a more "diffuse" attentional set, or one that weights lexical information more heavily, seems to be employed in the higher cognitive load conditions.

1.4 Signal factors and perceptual learning

A primary finding across the perceptual learning literature is that learning effects are only found on testing items that are similar to the exposure items. 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 must have some variance associated with them, and certain contexts can have increased degrees of variability. Kraljic et al. [23] 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. 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, but the coarticulatory effects for merging /s/ and /ʃ/ are much weaker in intervocalic position. 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. [24]. In addition to a continuum of /asi/ to /aʃi/, they also tested a continuum from /astri/ to /aʃtri/, and found comparable perceptual learning effects across both continua for those exposed to the intervocalic ambiguous sibilants, but no perceptual learning effects on either continua for the other condition, showing a context insentivity absent in other studies.

Sumner [40] investigated whether differences in presentation order in a perceptual learning experiment led to different learning effects of the /b/-/p/ category boundary for a native French speaker of English. The presentation order that showed the greatest perecptual learning effects was the one where tokens started out close to what a listener would expect for the categories (English-like voice onset time for /b/ and /p/) and shifted over the course of the experiment to what the speaker's actual categories were (French-like voice onset time for /b/ and /p/), despite the fact that this presentation order is not anything like what a listener would normally encounter when interacting with a non-native speaker of English. 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. One explanation for these results are that a small difference between the listener's expectations and the input provides more robust learning, and the future input uses the updated expectations for future input, allowing for greater shifts over time through smaller shifts per trial. In the exemplar model proposed by **?**], only input similar to the learned distribution is used for updating that distribution, and input too far away from learned distribution is discarded. ¡¡Double check the reference¿¿

1.5 Current contribution

Perceptual learning effects require two important aspects to be involved in the exposure phase. There must be some ambiguous acoustic aspect to be learned, and there must be an unambiguous link between the ambiguous acoustics and a sound category. This link can be provided through sensory information, such as in the visual domain [3], or it can be through linguistic information, such as the lexical status of the tokens [29]. Most of the literature within the domain of lexically-guided perceptual learning has been on the presence or absence of perceptual learning in the categorization phases, with manipulations to the categorization phase to test for generalization. The question that I am interested in is in the linking of ambiguous tokens to sound categories. Can manipulations in the exposure phase that reduce the reliability of the linking cause differences in perceptual learning?

In this dissertation, I will induce manipulations of lexical bias and attention in Experiments 1 and 2 (Chapter 2), and manipulations of sentence predictability and attention in Experiment 3 (Chapter 3). Lexical bias has been shown to affect phoneme categorization tasks [15], and can be manipulated by position of the ambiguous sound in the word and attention [30]. Sentence predictability, has likewise been found to affect phoneme categorization tasks similar to lexical bias [4], and can be manipulated by the preceding words in the sentence [18]. The interaction of sentence predictability with attention has not been explicitly studied, but the interaction should be similar to lexical bias, given findings that sentence predictability exerts similar effects on phone categorization as lexical bias [4]. In both of these studies, attention will be either diffuse across the word or focused on the speaker's phonetic characteristic to be learned. Attention to the speaker characteristic should lessen the acceptability of these productions as words as in previous work [30], leading to less perceptual learning when comparing these participants to those with a diffuse attentional set.

Chapter 2

Lexical decision

2.1 Motivation

The experiments in this chapter implement a standard lexically-guided perceptual learning experiment with manipulations to lexical bias and attention. In the perceptual learning literature, word endorsement rates during exposure are reported as high, such as no lower than 85% in some studies [34]. However, word endorsement rate of words containing ambiguous sounds varies critically as a result of the position of that sound, and thus the lexical bias present when a listener encounters it, and whether a listener's attention is directed to that sound [30]. Experiment 1 contains manipulations to these two factors across participants.

Additionally, 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 [40]. Words containing stimuli farther away from the target production are in general less likely to be endorsed as words, but similar effects of attention were found across word position [30]. 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.

2.2 Experiment 1

In these two experiments, listeners will be exposed to ambiguous productions of words containing a single instance of /s/, where the /s/ has been modified to sound more like /ʃ/ in a lexical decision task. In one group, the S-Initial group, the critical words will have an /s/ in the onset of the first syllable, like in *cement*, with no /ʃ/ neighbour, like *shement*. In the other group, the S-Final group, the critical words will have an /s/ in the onset of the final syllable, like in *tassel*, with no /ʃ/ neighbour like *tashel*. In addition, half of each group will be given instructions that the speaker has a ambiguous /s/ and to listen carefully, following Pitt and Szostak [30].

Given the difference in word response rates depending on position in the word, we would predict that listeners exposed to ambiguous sounds earlier in words would 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 former group would lead to a smaller perceptual learning effect as compared to the latter group.

2.2.1 Methodology

Participants

One hundred native speakers of English participated in the experiment and were compensated with either $10 CAD or course credit. They were recruited from the UBC student population. 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 forty English words and 100 nonwords that were phonologically legal in English 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 and half had an /s/ in the onset of the final

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/-initial	1.69 (0.85)	2.4 (0.59)
/s/-final	1.75 (1.11)	2.3 (0.47)
/ʃ/-initial	2.01 (1.17)	2.3 (0.48)
/ʃ/-final	1.60 (1.12)	2.4 (0.69)

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

syllable. 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 [6] 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 [19] in Matlab (The Mathworks, Inc.). Prior to morphing, the word and nonword versions were time aligned based on acoustic landmarks, like stop bursts, onset of F2, nasalization or frication, etc. All control items and filler words were processed and resynthesized by STRAIGHT to ensure a consistent quality across stimulus items.

Pretest

To determine which step of each continua would be used in exposure, a phonetic categorization experiment was conducted. Participants were presented with each step of each exposure word-nonword continuum and each categorization minimal pair continuum, resulting in 495 trials (40 exposure words plus five minimals pairs by 11 steps). The experiment was implemented in E-prime (cite). As half of the critical items had a sibilant in the middle of the word (onset of the final syllable), participants were asked to respond with word or non word rather than asking for the identity of the ambiguous sound, as in previous research [34].

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 this experiment was when the percentage of s-response dropped near 50%. A full list of steps chosen for each stimulus item is in the appendix. For the minimal pairs, six steps surrounding the 50% cross over point were selected for use in the phonetic categorization task. Due to experimenter error, the continuum for *seedling* was not included in the stimuli, so the chosen step was the average chosen step for the /s/-initial words. The average step chosen for /s/-initial words was 6.8 ($SD = 0.5$), and for /s/-final words the average step was 7.7 ($SD = 0.8$).

Procedure

Participants in the experimental conditions completed two tasks, an exposure task and a categorization task. The exposure task was a lexical decision task, where participants heard auditory stimuli and were instructed to respond with either ”word”

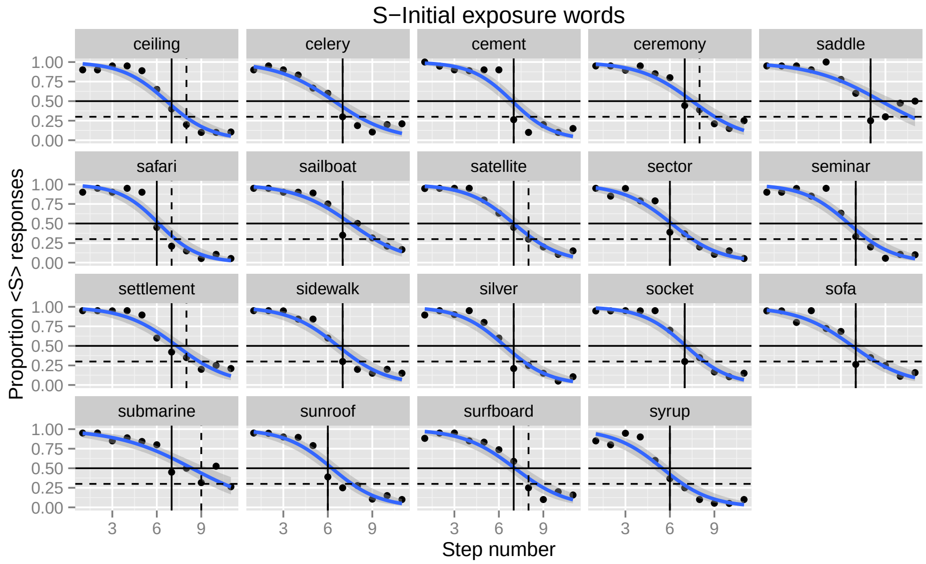


Figure 2.1: Proportion of word-responses for /s/-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line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

if they thought what they heard was a word or ”nonword” if they didn’t 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, but random otherwise, following Reinisch et al. [34].

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 (s vs sh). 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.

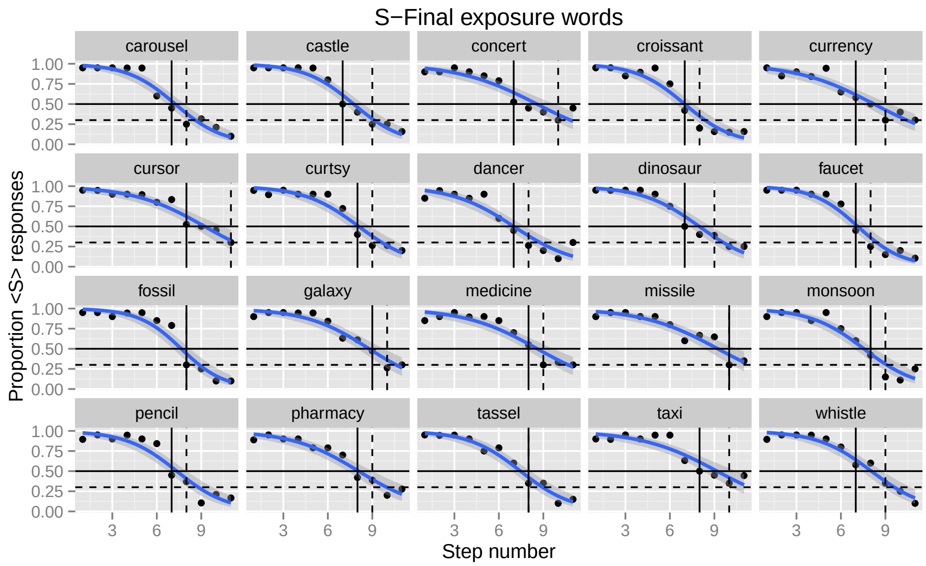


Figure 2.2: Proportion of word-responses for /s/-final exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line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

Participants were assigned to one of four conditions. Two of the conditions exposed participants to only criticall items that began with /s/, and the other two exposed them to only critical items that had an /s/ in the onset of the final syllable, giving a consistent 200 trials in all exposure phases with control and filler items shared across all participants. Additionally 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.

2.2.2 Results

Control experiment

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. A logistic mixed effects models was fit with Subject and

Continua as random effects and Step as a fixed effect with by-Subject and by-Item 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 ($\beta = -2.61, SE = 0.28, z = -9.1, p < 0.01$).

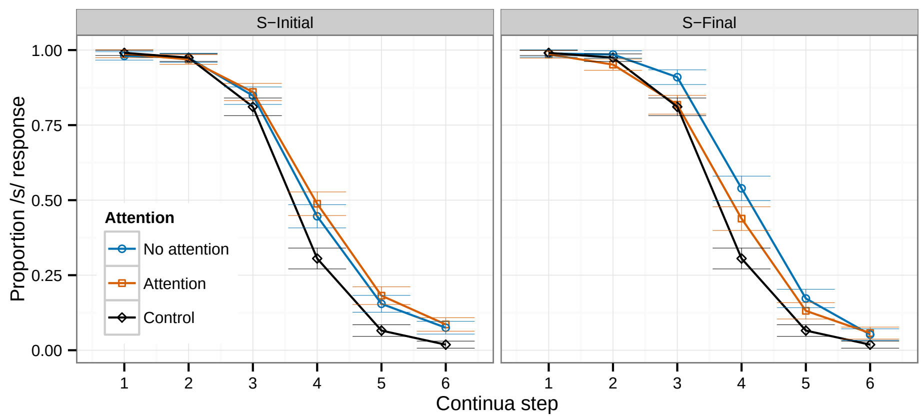
Exposure

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%. Word response rates for each of the four conditions did not differ significantly from each other, though S-Final/No Attention participants had a slightly higher average rate of 81% (SD= 17%) than the other conditions (S-Final/Attention: mean = 74%, SD = 18%; S-Initial/No Attention: mean = 74%, SD = 27%; S-Initial/Attention: mean = 76%, SD = 23%). A logistic mixed effects model with accuracy as the dependent variable was fit with fixed effects for trial type (Filler, S, SH), Attention (No Attention, Attention), Exposure Type (S-Initial, S-Final) and their interactions. The random effect structure was as maximally specified as possible with random effects for Subject and Word, and by-Subject random slopes for trial type and by-Word random slopes for Attention. The only fixed effects that were significant were a main effect of trial type for /s/ trials compared to filler trials ($\beta = -1.71, SE = 0.43, z = -3.97, p < 0.01$) and a main effect of Attention ($\beta = 0.76, SE = 0.38, z = 2.02, p = 0.04$). Trials containing an ambiguous /s/ were less likely to be responded to as a word, and participants instructed to pay attention to /s/ were more likely to correctly respond to words in general.

Categorization

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. Participants were excluded if their initial estimated cross over point for the continuum lay outside of the 6 steps presented (2 participants). 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.

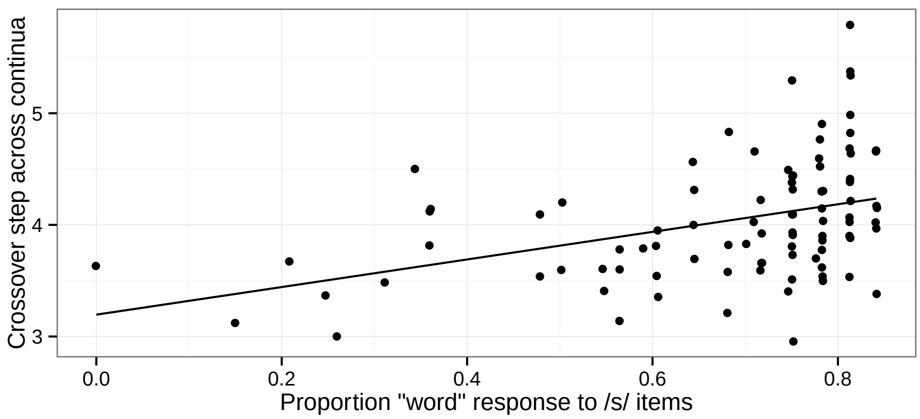
Figure 2.3: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 1. In the S-Final condition, participants in the Attention condition showed a larger perceptual learning effect than those in the No Attention condition. In the S-Initial condition, there were no differences in perceptual learning between the Attention conditions. Error bars represent 95% confidence intervals.



There was a significant effect for the intercept ($\beta = 0.83, SE = 0.31, z = 2.6, p < 0.01$), indicating that participants categorized more of the continua as /s/ in general. There was also a significant main effect of Step ($\beta = -2.10, SE = 0.20, z = -10.3, p < 0.01$), and a significant interaction between Exposure Type and Attention ($\beta = -0.93, SE = 0.43, z = -2.14, p = 0.03$). There was a marginal main effect of Exposure Type ($\beta = 0.58, SE = 0.30, z = 1.8, p = 0.06$).

These results are shown in Figure 2.3. The solid lines show the control participants' categorization function across the 6 steps of the continua. The error bars show within-subject 95% confidence intervals at each step. When exposed to ambiguous /s/ tokens in the first syllables of words, participants show a general expansion of the /s/ category, but no differences in behaviour if they are warned about ambiguous /s/ productions. 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/ than those who were warned of the speaker's ambiguous /s/ productions.

Figure 2.4: Correlation of crossover point in categorization with the proportion of word responses to critical items containing an ambiguous /s/ token.



As an individual predictor of participants' performance we took the proportion critical word endorsements and compared these values to the estimated crossover points. The crossover point was determined from the Subject random effect in the logistic mixed effects model [**?**]. There was a significant positive correlation between a participant's tolerance for the ambiguous exposure items and their crossover point on the continua ($r = 0.39, t(90) = 4, p < 0.01$), shown in Figure 2.4.

An ANOVA with cross-over point as the independent variable and word endorsement rate, Exposure Type, Attention and their interactions, found only a main effect of word endorsement rate ($F(1,89) = 17.82, p < 0.01$), suggesting that listeners in different conditions were not affected differently from one another.

2.2.3 Discussion

Perceptual learning effects in this experiment were robust across continua and experimental conditions, indicating the general automaticity of perceptual adaptation, but the degree of adaptation differed across conditions. Differences in attention only had an effect in the S-Final conditions, which replicates earlier findings that the effect of attention increased as the position of the ambiguous sound moved toward the end of the word [30]. The initial sound of a word is already a prominent position and listeners focus on the initial sounds to narrow the set of possible words they might be hearing. Later in a word, expectations for particular words given the preceding phonetic content would be greater, so directed attention on the phonetic detail shows a clear effect. That attention affected perceptual learning at all suggests that the adaptation is not wholly automatic and there is some degree of listener control.

The threshold for stimuli selection differed from that in previous studies. The closest study to this one used 70% of word responses in the pretest as the threshold for selection [34]. In that study, the lowest critical word endorsement rates in the exposure phase were 17 out of 20 (85%). In contrast, this experiment used 50% as the threshold and had correspondingly lower word endorsement rates (mean = 76%, sd = 22%). Interestingly, despite the lower word endorsement rates and the less canonical stimuli used, perceptual learning effects remained robust, which raises the question, can perceptual learning occur from stimuli that are farther from canonical productions than even the ones used in this experiment?

2.3 Experiment 2

Experiment 2 follows up on Experiment 1 by using stimuli that are farther from the canonical productions of the /s/ critical words. If the correlation from Experiment 1 holds, we expect to see lower word endorsement rates and lower (or perhaps non-existent) perceptual learning effects.

2.3.1 Methodology

This experiment followed an identical methodology 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/-final words the average step was 8.9 ($SD = 0.9$).

2.3.2 Results

Exposure

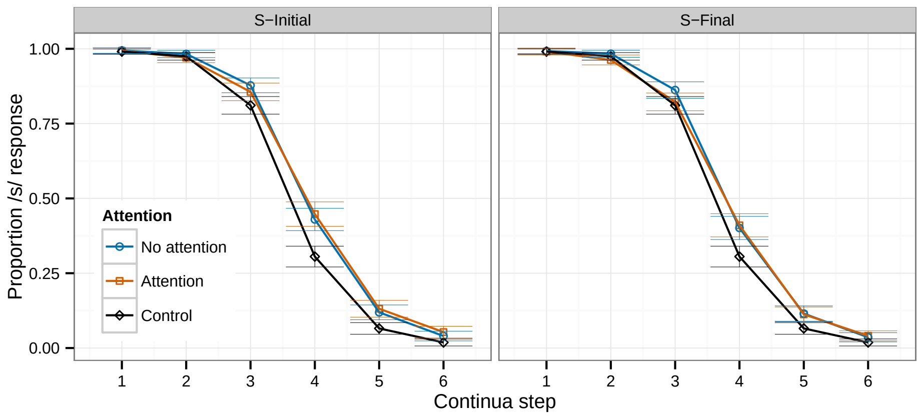
Trials with nonword stimuli and responses faster than 200 ms or slower than 2500 ms were excluded from analysis. Performance on the exposure task was high overall, with accuracy on filler trials averaging 92%. An ANOVA of critical word endorsement rates revealed a marginal effect of Exposure Type ($F(1,92) = 3.86, p = 0.05$), with participants in the S-Final conditions having lower word endorsement rates (S-Final/Attention: mean = 56%, sd = 30%; S-Final/No Attention: mean = 52%, sd = 25%) than participants in the S-Initial conditions (S-Initial/Attention: mean = 68%, sd = 25%; S-Initial/No Attention: mean = 61%, sd = 23%). A logistic mixed effects model with accuracy as the dependent variable was fit with fixed effects for trial type (Filler, S, SH), Attention (No Attention, Attention), Exposure Type (S-Initial, S-Final) and their interactions. The random effect structure was as maximally specified as possible with random effects for Subject and Word, and by-Subject random slopes for trial type and by-Word random slopes for Attention. The only fixed effect that was significant were a main effect of trial type for /s/ trials compared to filler trials ($\beta = -2.51, SE = 0.46, z = -5.35, p < 0.01$).

Categorization

Responses with reaction times less than 200 ms or greater than 2500 ms were excluded from analyses. Participants were excluded if their initial estimated cross over point for the continuum lay outside of the 6 steps presented (2 participants). 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.

There was a significant effect for the Intercept ($\beta = 1.01, SE = 0.38, z = 2.6, p <$

Figure 2.5: Proportion /s/ response along the 6 step continua as a function of Exposure Type and Attention in Experiment 2. Participants showed no significant differences across conditions. Error bars represent 95% confidence intervals.



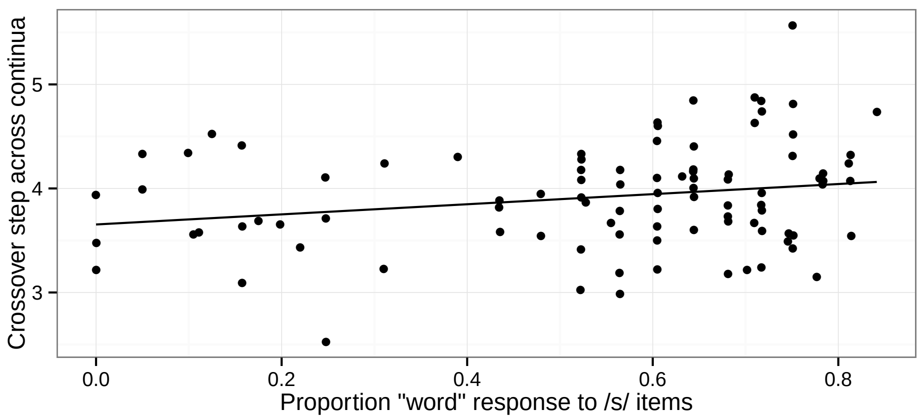
0.01), indicating that participants categorized more of the continua as /s/ in general. There was also a significant main effect of Step ($\beta = -2.67, SE = 0.23, z = -11.2, p < 0.01$). There were no other significant main effects or interactions, though an interaction between Step and Attention trended toward significant ($\beta = 0.35, SE = 0.21, z = 1.6, p = 0.09$).

As in Experiment 1, the proportion critical word endorsements was calculated for each subject and assessed for correlation with participants' crossover points. There was a significant positive correlation between a participant's tolerance for the ambiguous exposure items and their crossover point on the continua ($r = 0.22, t(92) = 2.25, p = 0.02$), shown in Figure 2.6.

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 fixed effects. In the lo-

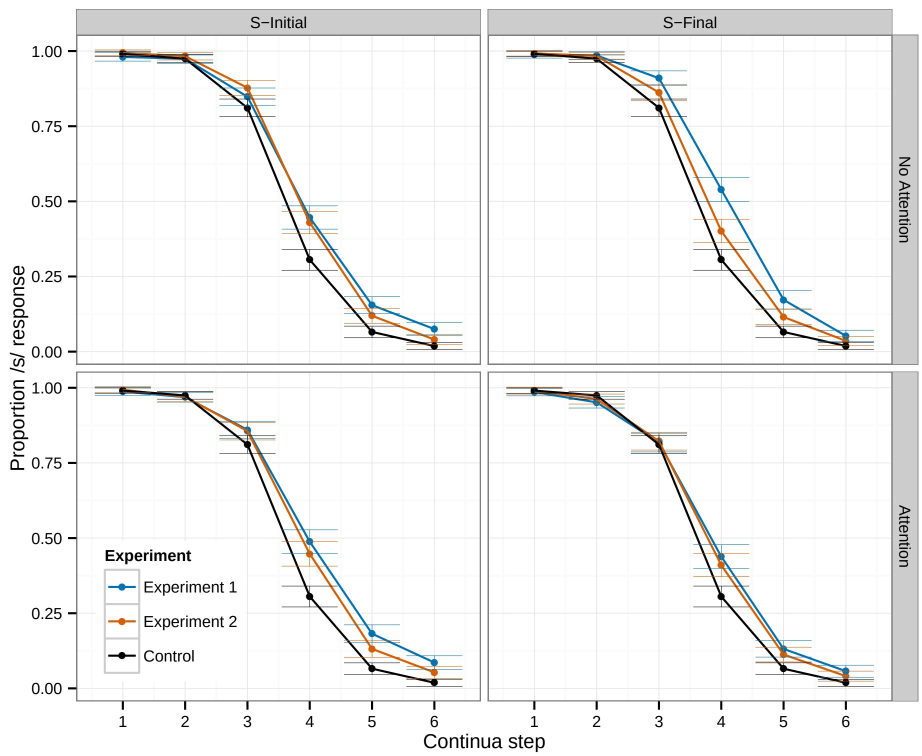
Figure 2.6: Correlation of crossover point in categorization with the proportion of word responses to critical items containing an ambiguous /s/ token in Experiment 2.



gistic mixed effects model, there was significant main effects for Intercept ($\beta = 1.00, SE = 0.36, z = 2.7, p < 0.01$) and Step ($\beta = -2.64, SE = 0.21, z = -12.1, p < 0.01$), and 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 2.7. The four-way interaction can be seen in S-Final/No Attention conditions across the two experiments, where Experiment 1 has a significant difference between the Attention and No Attention condition, but Experiment 1 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.

To see if there was a difference with the Experiment 1 in how word endorsement rates affected crossover points, the data was pooled for the two experiments. An ANOVA with cross-over point as the independent variable and word endorsement rate, Exposure Type, Attention, Experiment and their interactions, found a main effect of word endorsement rate ($F(1,185) = 21.82, p < 0.01$) and marginal

Figure 2.7: 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.



interaction between word endorsement rates and Experiment ($F(1,185) = 3.11, p = 0.07$).

2.5 General discussion

Perceptual learning was found across all experimental conditions. Attention only had effect on perceptual learning in the S-Final condition of Experiment 1.

Compared to Experiment 1, Experiment 2 had a weaker, yet still significant, correlation between critical word endorsement rates and crossover boundary points, suggesting 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 or position of the ambiguous sound in the word 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. Or perhaps more likely, the distance is inversely related to its goodness or plausibility that the ambiguous sound was indeed meant to be an /s/.

The correlation between word response rate in the exposure phase and the category boundary in categorization phase across both experiments raises two possible explanations. In a causal interpretation between exposure and categorization, as each ambiguous sound is linked to a word and a phonetic category, the distribution for that category (for that particular speaker) is updated. Participants who linked more of the ambiguous sound to the /s/ category updated their perceptual category for /s/ more. This explanation fits within a larger neo-generative model of spoken language processing [**?**]. 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, leading to greater degrees of perceptual adaptation. Individual differences in attention-switching control have previously been found to affect perceptual learning [39], which supports a non-causal interpretation as well.

The perceptual learning effects found in Experiment 1 and 2 appear to fall into two categories. For most of exposure conditions, participants showed a perceptual learning effect of similar magnitude. The pariticipants exposed to the ambiguous sounds with the most lexical bias and with no attention drawn to them, however, showed a significantly stronger perceptual learning effect. The disparity between these two cases and the lack of intermediate effects suggest that there may be two kinds of perceptual learning, one where lexical information is recruited and used and one which is wholly sensory. Any attention drawn to the ambiguous sound, either explicitly through instructions or through its position at the beginning or through its distance from the canonical production, discards or devalues the lexical connections leading to reduced generalization. This mode of perceptual learning perhaps exhibits more stimuli-specificity, like what is seen in psychophysics literature and in the lack of generalization in visually-guided perceptual learning in speech percetion [35]. The wholly sensory perceptual learning might only show effects for the particular words that were in the exposure phase, as the participants exposed to the ambiguous /s/ at the beginnings of words did not show a greater perceptual learning effect on the test continua which had a sibilant at the beginning.

It may be the case the influence of attention on the linguistic bias used in perceptual learning is in fact gradient, but the scales of expectation used in the experiments in this chapter were too small to see statistical effects. To test whether attention exerts a categorical influence or a more gradient one, the experiment in the following chapter increases the linguistic expectations of the target items through semantic predictability in addition to lexical bias.

Chapter 3

Cross-modal word identification

3.1 Motivation

In Chapter 2, listeners showed greater perceptual learning effects when the lexical bias was greater. However, this relationship was not found when attention was drawn to the ambiguous sound and when the ambiguous sound was farther from the canonical production. The experiment in this chapter seeks to increase the linguistic expectations for the words containing the ambiguous productions to test whether attention categorically removes the effect of linguistic expectation, or merely moderates its effect. To this end, semantic predictability is used to boost linguistic expectations in conjunction with lexical bias.

Semantic predictability has mostly been used in production studies, where it has been found to have an effect on vowel realization and duration independent of lexical factors such as frequency and neighbourhood density [9, 37]. In speech perception work, semantic predictability has been found to have an effect on phoneme categorization similar to that of lexical bias [4], and increased intelligibility in noise, particularly for native speakers [18**? ? ?** , and others].

As in previous work, two levels of semantic predictability are used in this experiment. Highly predictable sentences are ones where the final word is constrained to only a few words. Unpredictable sentences are ones where the possible completions are numerous, and any completions given are more guesses than anything else. All target words and the threshold for ambiguity are taken from Experiment

1 and have the highest lexical bias available. When a listener is exposed to these words in unpredictable sentences, they should show perceptual learning effects equivalent to those participants in the S-Final condition of Experiment 1.

The exposure task in this experiment differs from lexical decision task used in the previous chapter. Instead, participants identify the picture that matches the final word in the sentence, and all trials involve a real word. The attentional set for this task, where participants make a decision about the identity of a specific word, may be different from the previous experiment, where they make a decision about the lexicality of a word, but the decision is about the word in both instances.

If linguistic expectations are cumulative, we should expect to see a larger perceptual learning effect for participants exposed to ambiguous sounds in words that are highly predictable. Additionally, if attention does exert a categorical effect on perceptual learning, we should see a similar size of perceptual learning in this experiment in the attention conditions as in the previous experiments.

3.2 Methodology

3.2.1 Participants

One hundred native speakers of English (mean age ??, range ??-??) participated in the experiment and were compensated with either $10 CAD or course credit. They were recruited from the UBC student population. 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.

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 somtimes ambiguous, and to listen carefully to ensure correct responses. Participants were native North American English speakers with no reported speech or hearing disorders.

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 [18], Unpredictive sentences were written with the final word in mind with a variety of sentence structures, and 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 [19]; however, only the final word in sentence was morphed. For all steps, 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 2.

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, 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 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 (cite), and were sanitized by removing miscellaneous keystrokes recorded by E-Prime, 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 from the Speech in Context lab 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 in the first pretest.

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 two tasks, 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 presented with two pictures on the screen. Their task was to select the picture on the screen that corresponded to the final word in the sentence they heard. As in Experiments 1 and 2, the order was pseudorandom, with the same constraints.

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

3.3 Results

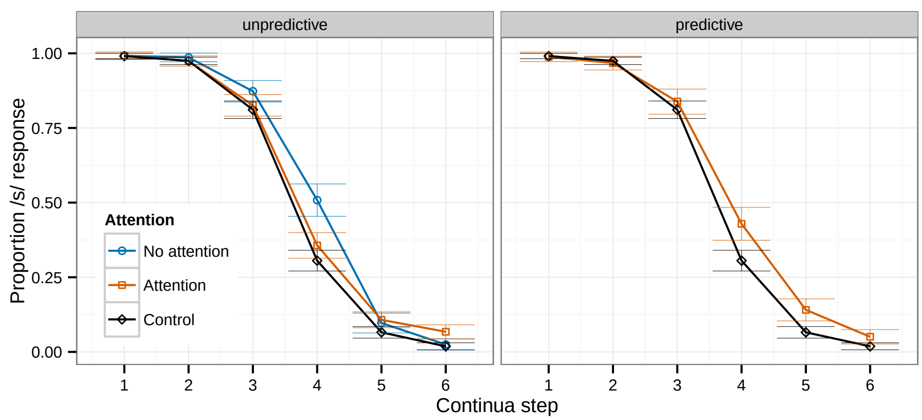
3.3.1 Exposure

Performance in the task was high, with accuracy at ceiling.

3.3.2 Categorization

3.4 Discussion

Figure 3.1: 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.



Chapter 4

Conclusions

4.1 Effect of increased linguistic expectations

Bias for a word was manipulated in two ways, through position of the ambiguous sound in the word and through sentential manipulations. The conditions that correspond most closely to those in other lexically-guided perceptual learning experiments [21, 29] are those in the No Attention condition of Experiment 1. Under those conditions, there is a clear effect of lexical bias, such that listeners exposed to ambiguous stimuli with greater lexical bias update their speaker-specific category for that sound more.

4.2 Attentional control of perceptual learning

The findings of Experiment 1 suggest also that perceptual learning is not a wholly automatic process. 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 witheld. Attention on the ambiguous sound equalized the perceptual learning effects across lexical bias. However, in Experiment 2, there is no such effect of attention, suggesting that the ambiguous sounds that were farther away from the canonical production drew the listener's attention to those sounds.

One question raised by the current results is whether attention always decreases 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? I would predict that there would be a greater tendency to endorse ambiguous productions as words when the attention is positive as compared to both no attention and negative attention, and would therefore lead to greater perceptual learning effects. This prediction will be tested in future work.

4.3 Specificity versus generalization in perceptual learning

The results of this dissertation speak to dichotomy between specificity and generalization found in perceptual learning work. In Experiment 1, greater perceptual learning was shown by participants exposed to ambiguous sounds later in the words, not to those at the beginnings of words. And yet, the testing continua consisted of stimuli with the sibilant at the beginnings of words, which are more similar to the words beginning with the ambiguous sound.

The effect of attention in Experiments 1 and 2 suggest that when speakers are focused on the signal, or have their attention drawn to the signal, they are less likely to generalize across forms. When attention on the signal is not present, the available lexical connections can assist in greater generalization to new forms. These different modes of perceptual learning may account for some differences that have been observed in the literature. While visually-guided perceptual learning shows comparable effects to lexically-guided perceptual learning [41], lexically-guided perceptual learning is more likely to be expanded to new contexts [23, 29], though with some restrictions [28], than visually-guided perceptual learning [35].

4.4 Distance to canonical production

In Experiment 2, there was no effect of attention or lexical bias on perceptual learning, with a stable 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 similar pattern as for listeners in Experiment 1 in the Attention condition. The second poential 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 [**?**]. 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 [40] where the most perceptual learning was found when the categories begin like the listener expects and gradually shift toward the speaker's actual boundaries over the course of exposure.

These two explanations could be tested if the predictions for positive attention are borne out. If positive attention does increase word endorsement rate as predicted, it should have an effect on exposure items farther away from canonical productions as well. If those productions do have a weaker effect on updating listener categories, then we should see a weaker perceptual learning effect across ambiguous stimuli types despite consistent word endorsement rates.

An additional way to test the second explanation would be to implement a paradigm similar to that used in Sumner [40]. It is unlikely that the non-nativeness of the speaker is the primary reason for the findings. While a listener might expect a non-native speaker to produce speech farther from the expected distributions of native speakers, the gradual transition from English-like stops to French-like stops is at least as unrealistic as the other conditions in Sumner [40]. One study that included a manipulation to the order of presentation of ambiguous and unambiguous stimuli is Kraljic et al. [24], which found that when speakers are presented with 10 ambiguous /s/-/ʃ/ stimuli and 10 nonambiguous /s/ stimuli in different orders, perceptual learning differs. When the ambiguous stimuli are presented first followed by the nonambiguous ones, listeners show a perceptual learning effect, but no perceptual learning effect is found when the order is reversed. They argue that listeners only adapt their categories if the ambiguous stimuli are reliable evidence. So in the case where participants heard the ambiguous stimuli after the nonambiguous ones, the ambiguous stimuli would not be interpreted as reliable in light of the previous tokens that the listener had heard. Because the difference between the initial stimuli and the later ones in Kraljic et al. [24] was a categorical one, it may have drawn the attention of the listener like the stimuli used in Experiment 2. Perhaps a more the gradient change as in Sumner [40] could keep the differences beneath the attentional threshold of the listener and thus induce greater perceptual learning. If participants had a larger perceptual learning effect, as compared to the participants in Experiment 1, after exposure to a presentation order that began as more canonical productions and then gradually shifted to productions for the 50% ambiguous point, then the predictions of **?**] would be further supported.

This disseratation investigated the interaction between linguistic, attentional and signal factors influencing perceptual learning in speech perception. Increased linguistic bias resulted in larger perceptual learning effects. However, perceptual learning was modulated by the attention of the listener and the propterties of the signal. If a listener was warned of the ambiguous sound, the effect of increased linguistic bias on perceptual learning disappeared. Likewise, if the ambiguous sound was farther from the intended target, there was no effect of linguistic bias or attention on perceptual learning. These results suggest that the degree to which listeners perceptually adapt to a new speaker is under their control to a degree. All participants as a whole exhibited a perceptual learning effect, indicating that perceptual learning is to some extent an automatic process.

Bibliography

[1] M. Ahissar and S. Hochstein. Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12):5718–22, 1993. ISSN 0027-8424. doi:10.1073/pnas.90.12.5718. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46793&tool=pmcentrez&rendertype=abstract. → pages 9

[2] W. F. Bacon and H. E. Egeth. Overriding stimulus-driven attentional capture. *Perception & psychophysics*, 55(5):485–496, 1994. ISSN 0031-5117. → pages 9

[3] P. Bertelson, J. Vroomen, and B. De Gelder. Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect. *Psychological Science*, 14(6):592–597, 2003. → pages 3, 12

[4] S. Borsky, B. Tuller, and L. P. Shapiro. "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, 1998. URL http://www.ncbi.nlm.nih.gov/pubmed/9604360. → pages 8, 13, 30

[5] A. R. Bradlow and T. Bent. Perceptual adaptation to non-native speech. *Cognition*, 106(2):707–729, 2008. → pages 1, 7

[6] M. Brysbaert and B. New. 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, 2009. ISSN 1554-3528. → pages 16

[7] E. Clare. Applying phonological knowledge to phonetic accommodation, 2014. → pages 7

[8] A. Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and brain sciences*, 36(3):181–204,

2013. ISSN 1469-1825. URL
http://www.ncbi.nlm.nih.gov/pubmed/23663408. → pages 4

[9] C. G. Clopper and J. B. Pierrehumbert. Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, 124(3):1682–1688, Sept. 2008. ISSN 1520-8524. doi:10.1121/1.2953322. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2676620&tool=pmcentrez&rendertype=abstract. → pages 7, 8, 9, 30

[10] C. Connine. Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 538:527–538, 1987. URL http://www.sciencedirect.com/science/article/pii/0749596X87901380. → pages 8

[11] C. M. Connine and C. Clifton. Interactive use of lexical information in speech perception. *Journal of experimental psychology. Human perception and performance*, 13(2):291–299, 1987. ISSN 0096-1523. → pages 6, 8, 9

[12] P. D. Eimas, W. E. Cooper, and J. D. Corbit. Some properties of linguistic feature detectors, 1973. ISSN 0031-5117. → pages 4

[13] F. Eisner and J. M. McQueen. The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2):224–238, 2005. → pages 4

[14] F. Eisner and J. M. McQueen. Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4):1950, 2006. ISSN 00014966. doi:10.1121/1.2178721. URL http://scitation.aip.org/content/asa/journal/jasa/119/4/10.1121/1.2178721. → pages 3

[15] W. F. Ganong. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1): 110–125, 1980. URL http://www.ncbi.nlm.nih.gov/pubmed/6444985. → pages 5, 13

[16] E. J. GIBSON. Improvement in perceptual judgments as a function of controlled practice or training. *Psychological bulletin*, 50(6):401–431, 1953. → pages 3

[17] C. Gilbert, M. Sigman, and R. Crist. The neural basis of perceptual learning. *Neuron*, 31:681–697, 2001. ISSN 0896-6273. doi:10.1016/S0896-6273(01)00424-X. URL http://www.sciencedirect.com/science/article/pii/S089662730100424X. → pages 4

[18] D. Kalikow, K. Stevens, and L. Elliott. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *... Journal of the Acoustical Society of ...*, 61(5), 1977. URL http://link.aip.org/link/?JASMAN/61/1337/1. → pages 7, 8, 13, 30, 32

[19] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno. 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, 2008. → pages 17, 32

[20] D. Kleinschmidt and T. F. Jaeger. A Bayesian belief updating model of phonetic recalibration and selective adaptation. *Science*, (June):10–19, 2011. URL http://www.aclweb.org/anthology-new/W/W11/W11-06.pdf#page=20. → pages 4

[21] T. Kraljic and A. G. Samuel. Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, 51(2):141–78, Sept. 2005. ISSN 0010-0285. doi:10.1016/j.cogpsych.2005.05.001. URL http://www.ncbi.nlm.nih.gov/pubmed/16095588. → pages 3, 4, 7, 36

[22] T. Kraljic and A. G. Samuel. Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1):1–15, Jan. 2007. ISSN 0749596X. doi:10.1016/j.jml.2006.07.010. URL http://linkinghub.elsevier.com/retrieve/pii/S0749596X06000842. → pages 4

[23] T. Kraljic, S. E. Brennan, and A. G. Samuel. Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, 107(1):54–81, Apr. 2008. ISSN 0010-0277. doi:10.1016/j.cognition.2007.07.013. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2375975&tool=pmcentrez&rendertype=abstract. → pages 7, 11, 37

[24] T. Kraljic, A. G. Samuel, and S. E. Brennan. First impressions and last resorts: how listeners adjust to speaker variability. *Psychological science*, 19 (4):332–8, Apr. 2008. ISSN 0956-7976.

doi:10.1111/j.1467-9280.2008.02090.x. URL http://www.ncbi.nlm.nih.gov/pubmed/18399885. → pages 7, 12, 38

[25] A. B. Leber and H. E. Egeth. Attention on autopilot: Past experience and attentional set, 2006. ISSN 1350-6285. → pages 10

[26] P. Lieberman. Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech. *Language and Speech*, 6(3):172 –187, 1963. URL http://las.sagepub.com/content/6/3/172.abstract. → pages 8

[27] S. L. Mattys and L. Wiget. Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160, 2011. → pages 11

[28] H. Mitterer, O. Scharenborg, and J. M. McQueen. Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2):356–361, 2013. → pages 4, 37

[29] D. Norris, J. M. McQueen, and A. Cutler. Perceptual learning in speech, 2003. → pages 2, 3, 4, 5, 6, 10, 12, 36, 37

[30] M. Pitt and C. Szostak. A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, (April 2013):37–41, 2012. URL http://www.tandfonline.com/doi/abs/10.1080/01690965.2011.619370. → pages 6, 7, 10, 11, 13, 14, 15, 23

[31] M. A. Pitt and A. G. Samuel. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of experimental psychology. Human perception and performance*, 19(4):699–725, 1993. ISSN 0096-1523. → pages 6

[32] M. A. Pitt and A. G. Samuel. Word length and lexical activation: longer is better. *Journal of experimental psychology. Human perception and performance*, 32(5):1120–1135, 2006. ISSN 0096-1523. → pages 6

[33] E. Reinisch and L. L. Holt. Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. *Journal of experimental psychology. Human perception and performance*, 40(2):539–555, 2013. ISSN 1939-1277. URL http://www.ncbi.nlm.nih.gov/pubmed/24059846. → pages 4

[34] E. Reinisch, A. Weber, and H. Mitterer. Listeners retune phoneme categories across languages. *Journal of experimental psychology. Human perception*

and performance, 39(1):75–86, Feb. 2013. ISSN 1939-1277. doi:10.1037/a0027979. URL http://www.ncbi.nlm.nih.gov/pubmed/22545600. → pages 4, 14, 17, 18, 23

[35] E. Reinisch, D. R. Wozny, H. Mitterer, and L. L. Holt. Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45:91–105, 2014. ISSN 00954470. doi:10.1016/j.wocn.2014.04.002. URL http://linkinghub.elsevier.com/retrieve/pii/S009544701400045X. → pages 4, 29, 37

[36] A. G. Samuel. Red herring detectors and speech perception: in defense of selective adaptation. *Cognitive psychology*, 18(4):452–499, 1986. ISSN 00100285. → pages 4

[37] R. Scarborough. Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory phonology*, pages 575–604, 2010. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Lexical+and+contextual+predictability:+confluent+effects+on+the+production+of+vowels#0. → pages 7, 8, 9, 30

[38] O. Scharenborg and E. Janse. Comparing lexically guided perceptual learning in younger and older listeners. *Attention, perception & psychophysics*, 75(3):525–36, 2013. ISSN 1943-393X. URL http://www.ncbi.nlm.nih.gov/pubmed/23354594. → pages 10

[39] O. Scharenborg, A. Weber, and E. Janse. The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception, & Psychophysics*, pages 1–15, 2014. → pages 10, 28

[40] M. Sumner. The role of variation in the perception of accented speech. *Cognition*, 119(1):131–136, 2011. ISSN 0010-0277. doi:10.1016/j.cognition.2010.10.018. URL http://dx.doi.org/10.1016/j.cognition.2010.10.018. → pages 12, 14, 38

[41] S. van Linden and J. Vroomen. Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of experimental psychology. Human perception and performance*, 33(6):1483–1494, 2007. ISSN 0096-1523. → pages 3, 37

[42] L. van Noorden and J. F. Schouten. *Temporal Coherence in the Perception of Tone Sequences*. PhD thesis, 1975. → pages 10

[43] J. Vroomen, S. van Linden, B. de Gelder, and P. Bertelson. Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3):572–577, 2007. → pages 4

[44] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature reviews. Neuroscience*, 5(6): 495–501, 2004. → pages 9