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.

Don’t forget your parents or loved ones.

You may wish to acknowledge your funding sources.

Chapter 1

Introduction

Stuff I have introduced

Chapter 2

Background

Listeners of a language are faced with a large degree of 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 harder to separate in acoustic domains. Despite this, listeners maintain a large degree of perceptual constancy, through speaker normalization and perceptual learning. Ambiguous vowel sounds can be mapped to different categories depending on the vowels of the preceding context [18], where the previously-unheard ambiguous tokens is normalized to the vowel space present in the context. Perceptual learning refers to the updating of distributions corresponding to a sound category for a particular speaker.

2.1 Perceptual learning

Norris et al. [22] began the investigation into perceptual learning in speech. Previous work has demonstrated perceptual learning in the visual domain (Color hue reference?) and (I think auditory as well). Norris et al. [22] exposed Dutch listeners to a fricative halfway between /s/ and /f/ at the ends of words and nonwords and tested 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. [1], 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. These categorization tests showed clear perceptual recalibration effects, such that the participant was 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 [30] compared the perceptual recalibration effects from the visual lipread paradigm [1] to the standard lexically-guided perceptual learning paradigm [22]. 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 [8, 13].

2.2 Retention

Kraljic and Samuel [13] looked at the strength of the perceptual learning effect across various types of unlearning tasks. Native English listeners were exposed to ambiguous fricatives in between /s/ and /ʃ/ in words that biased interpretation toward either /s/ or /ʃ/ in a lexical decision task either spoken by a female voice or

a male voice. Participants that completed an /s/-/ʃ/ categorization task immediately following exposure and also those that had a 25 minute visua Unlearning task between exposure and categorization showed a large perceptual learning effect. Additionally, participants that had an Unlearning task involving auditory input from the same speaker as they were exposed to showed attenuated perceptual learning effects when the speech contained correct versions of the ambiguous fricatives, i.e. correct /s/ tokens when they were exposed to the ambiguous sibilant in all /s/ words in the exposure task. The perceptual learning effect in this condition was almost eliminated. Participants that had an Unlearning task involving auditory input that did not have any /s/ or /ʃ/ tokens showed perceptual learning effects comparable to those with no Unlearning task. When the Unlearning task involved auditory input from a different speaker than exposure, perceptual learning effects were as robust as when there was no Unlearning task.

Eisner and McQueen [8] looked at the stability of perceptual learning. Dutch participants first completed a categorization task for a continuum of /s/ to /f/ and then listened to a story that contained ambiguous fricative between /s/ and /f/, with one version having the ambiguous fricative in /s/ contexts and the other having it in /f/ contexts. Immediately following, they completed another /s/ to /f/ categorization task. Finally, 12 hours later, they completed the categorization task once again. Half the participants started at 9 am and completed the final categorization task at 9 pm the same day, and the other half started at 9 pm and completed the final categorization at 9 am the following day. Perceptual learning effects were found in both categorization tasks following exposure, with no significant differences between the two times. The perceptual learning effects found were robust across time and across intervening language exposure, as the those tested at 9 am and 9 pm likely had many interactions with other people in between.

Perceptual learning effects are robust across time. Given no other interactions with a given speaker, a listener's behaviour for that speaker's speech will remain constant over the course of a day, and probably longer. However, the perceptual system remains plastic. If a listener is exposed to a speaker trait in the first interaction, but the trait disappears in later interactions, the perceptual system will reattenuate to the newer evidence.

2.3 Generalization

2.3.1 Generalizing to new speakers

Kraljic and Samuel [15] looked at different sound contrasts and the ability of participants to generalize across speakers. The first contrast was voicing between /t/ and /d/. When exposed to ambiguous stops between /t/ and /d/ corresponding to /d/ for one speaker, and ambiguous stops correpsonding to /t/ for another, perceptual learning effects corresponded to the most recent exposure, regardless of whether the test voice was the same as that recent exposure voice. The second contrast they used was sibilants of /s/ and /ʃ/. For this contrast, they found the speaker-specificity effect found in the literature, where recency of the exposure did not affect the perceptual learning for a particular voice. They argue that the differences between these two contrasts lies in the relative indexical information carried by the sibilants and the lack of much indexical information carried in the stops.

Eisner and McQueen [7] looked at what manipulations could cause perceptual learning could generalize to additional talkers. Using a Dutch lexical decision task with ambiguous fricatives in between /s/ and /f/, they tested categorization on a contniuum between /s/ and /f/ from two talkers. When exposed to one talker's ambiguous fricative and tested on another speaker's continuum, no perceptual learning effect was found. There were two conditions where the fricatives in the exposure and categorization came from the same speaker, but the speakers of the carrier speech. In one, the ambiguous fricative from the categorization speaker's produtions was spliced into the exposure tokens, and in the other, the exposure talker's /s/ to /f/ continuum were spliced with vowels from the categorization talker. In these two conditions, they did find a perceptual learning effect across speakers, despite reports by the participants of hearing different speakers, suggesting a sound-specificity effect in addition (or potentially causing) speaker-specificity effects.

Reinisch et al. [27] looked at perceptual learning of speaker characteristics across languages. When exposed to a native Dutch speaker speaking English with ambiguous fricatives between /s/ and /f/, participants show a perceptual learning effect when categorizing Dutch minimal pairs differing only in whether one sound is an /f/ or /s/ spoken by the same speaker. When exposed to the same native Dutch

speaker speaking Dutch words and nonwords with the same ambiguous fricative, native Dutch listeners and L2 Dutch listeners (native German speakers) show comparable perceptual learning effects.

Reinisch and Holt [26] Holt exposed native English listeners to a female Dutch-accented voice speaking words and nonwords, where words ending in /s/ or /f/ instead ended either in a natural token or an ambiguous fricative between /s/ and /f/. Listeners showed a perceptual learning effect for the exposure talker as well as another female Dutch-accented voice, showing cross-speaker generalization. However, when categorizing a continuum from /s/ to /f/ of a male Dutch-accented voice, listeners did not generalize the perceptual learning effect, but there was a slight perceptual learning effect for the first block of categorization tokens for the male voice that disappeared in later blocks. Pretests of the continua revealed that the male voice continuum was heavily biased towards /s/, with steps 1-4 of the continuum consistently heard as /s/. For the exposure female voice, Steps 1-4 were already ambiguous between /s/ and /f/, with the proportion /s/ response less than 60% for all steps. The new female voice was responded to similar to the exposure female voice, but with less ambiguity for the initial step. When the continua for the exposure female voice and the male voice were more closely matched by removing the first four steps of the male voice and removing 4 steps along the continuum of the female voice, generalization of the perceptual learning effects matched those for the new female voice. These findings suggest that the perceptual system leverages known distributions that may not be speaker specific. When new tokens fall outside that known distribution, as in the male voice categorization, perceptual learning for that voice occurred using the relatively abundant unambiguous examples for that speaker to adjust the perceptual system for that talker.

Kraljic and Samuel [13], in addition to the findings reported in in Section 2.2, also found an asymmetry in how listeners generalized perceptual learning. Across all of their experiments, categorization of the same voice as exposure produced perceptual learning. When the continuum from /s/ to /ʃ/ from the male voice was categorized following exposure to the female voice, perceptual learning effects were consistently present. When the voice of the exposure and categorization was reversed, the consistency of perceptual learning effects disappeared. They performed an acoustic analysis of the spectral means for the categorization items and

the exposure items for each talker. The spectral means of the voices for the continuum steps were well separated, but the exposure items were much closer together. The spectral mean of the ambiguous sibilant for the female voice fell in between the ranges of the female continuum steps and the male continuum steps, but the spectral mean exposure ambiguous sibilant for the male voice was squarely in the range of the male continuum steps. The asymmetry of speaker generalization can then be attributed to the female exposure ambiguous items being acoustically (and therefore perceptually) similar enough to generalize the perceptual learning for the female voice to the male voice.

From these studies, the clearest thread through all of them is the degree of acoustic similarity or perceptual similarity between the exposure stimuli that prompts perceptual learning and whatever continuum the listeners are tested on. Stops and fricative do not actually appear to behave categorically different, per se, but their differences in behaviour can be attributed to the fact that stops are acoustically more similar to each other in general than fricatives are.

2.3.2 Generalizing to other categories and contexts

Kraljic and Samuel [14] looked at two sound contrasts (/d/-/t/ and /b/-/p/) that shared a feature, namely voicing. Participants were exposed to ambiguous in between /d/ and /t/ in English words and then tested on two continua with two different voices. The categorization continuum that was presented first, from /b/ to /p/, showed the stronger perceptual learning effects than the second continuum from /d/ to /t/, even though participants were exposed to ambiguous tokens between /d/ and /t/. Both old voices and new voices for the categorization task produced perceptual learning effects.

Mitterer et al. [21] exposed participants to either words ending with /r/ or /l/ or to words beginning with /r/ or /l/, and tested each group's perceptual learning behaviour on continua from /r/ to /l/ in both initial and final positions. If listeners were exposed to an ambiguous sound between /r/ and /l/ at the ends of words, they showed a perceptual learning effect for the continuum that ended with an /r/ or /l/, but not for the continuum that began with an /r/ or an /l/. The inverse behaviour was observed for the group that was exposed to an ambiguous sound between /r/

and /l/ at the beginnings of words. They argue that the category being affected is not a context-insensitive phoneme, but rather a context sensitive allophone. Jesse and McQueen [10] found that listeners did indeed generalize across positions, but their stimuli were the fricatives /s/ and /f/, and the same physical fricatives were the same across all positions.

Reinisch et al. [28] utilized a visually-guided recalibration paradigm from Bertelson et al. [1] to test whether acoustic cues to place of articulation could be recalibrated with unambiguous visual feedback. In Experiment 1, they trained listeners by exposing one group to the maximally ambiguous token between /aba/ and /ada/ with either a video of /ada/ or /aba/ and another group to the maximally ambiguous token between /ibi/ and /idi/ with either a video of /ibi/ or /idi/. In vowel context of /a/, the consonant portion of the token was silenced so that only formant cues to place of articulation would be available. In the vowel context of /i/, the consonant part was not silenced, but since formant transitions in the /i/ context are the same for labials and dentals, the only cues to place of articulation would be in the consonant itself. Afterwards, when they categorized auditory only continua of /aba/ to /ada/ and /ibi/ to /idi/ with the same silencing of the consonant in the /a/ context. They found a recalibration effect for the continuum participants were exposed to, but no generalization effect to the novel continuum. Experiment 2 followed the same setup as Experiment 1, with the same /aba/ to /ada/ as in Experiment 1, but replaced the /ibi/ to /idi/ continuum with a /ama/ to /ana/ continuum. As in Experiment 1, clear recalibration effects were found for the continuum participants were exposed to, but not for the novel continuum. Experiment 3 followed the same procedure but used a /ubu/ to /udu/ continuum instead of an /ibi/ to /idi/ continuum. The /u/ context provides formant transition cues to place of articulation, so the consonant was silenced for this continuum as well. As before, there were perceptual recalibration effects for the continuum participants were exposed to, but not to the novel continuum. Across all of these, there is a context specifity effect.

Kraljic et al. [16] 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/ of-

ten 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. [17]. 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.

2.4 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 bottom-up information, such as in the visual domain [1], or it can be top-down information, such as the lexical status of the tokens [22]. 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 significant differences in perceptual learning?

In this dissertation, I will induce manipulations of lexical bias and attention in Experiments 1 and 2, and manipulations of sentence predictability and attention in Experiment 3. Lexical bias has been shown to affect phoneme categorization tasks [9], and can be manipulated by position of the ambiguous sound in the word and attention [23]. Sentence predictability, has likewise been found to affect phoneme categorization tasks similar to lexical bias [2], and can be manipulated by the preceding words in the sentence [11]. The interaction of sentence predictability with attention has not been explicitly studied, but the interaction should be similar to

lexical bias.

Chapter 3

Lexical decision

3.1 Motivation

3.1.1 Lexical bias

Lexical bias is the primary way through which perceptual learning occurs in the experimental literature. Lexical bias, also known as the Ganong Effect, refers to the tendency for listeners to interpret a speaker's production as a 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 [9]. 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.

Studies looking at lexical bias 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.

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 [25]. 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 [25] also found that lexical bias from monosyllabic words show greater response to experimental manipulations than trisyllabic words.

Pitt and Szostak [23] 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 [24], they found that, for monosyllabic word frames, token-final targets produce more stable results than token-initial targets.

Pitt and Szostak [23] 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.

Lexical bias is affected by processing conditions. In their first experiment, Mattys and Wiget [20] found that lexical bias has a larger effect when listeners have more cognitive load than when they have no cognitive load from concurrent tasks. In their second experiment, the same cognitive load and non-cognitive load conditions were present, but listeners had to respond to the phoneme identification task within 1500 ms. In this experiment, the lexical bias across cognitive load conditions was identical and similar to the cognitive load condition of experiment 1. In experiment 3, listeners could only respond *after* 1500 ms, and the effect of cognitive load re-emerged with similar magnitude to experiment 1. In the remainder of their experiments, they examined whether cognitive load increased word activation or decreased sensory information, the two primary possible explanations for

the pattern of results found in experiment 1. They found that semantic priming was not significantly different across cognitive loads, but fine-grained discrimination suffered under cognitive load conditions. Thus, the effect of cognitive load observed is a function of impoverished fine detail leading to a greater reliance on lexical information. However, the timecourse for the incorporation of these different sources of information is interesting, with lexical information playing a larger role earlier on in processing and sensory signal information being incoporated later on. Earlier responses are more likely to be lexically biased than later responses.

3.2 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 [23].

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.

3.2.1 Methodology

Participants

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

Table 3.1: Mean and standard deviations for frequencies 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)

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 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 or ʤ/. Filler words and nonwords did not contain any sibilants. Frequencies and number of syllables across item types are in Table 3.1

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 [3] for the /s/ word, and two had higher LFPM for the /ʃ/ word, as shown in Table 3.2.

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

Table 3.2: Frequencies 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

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 [12] 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 [27].

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

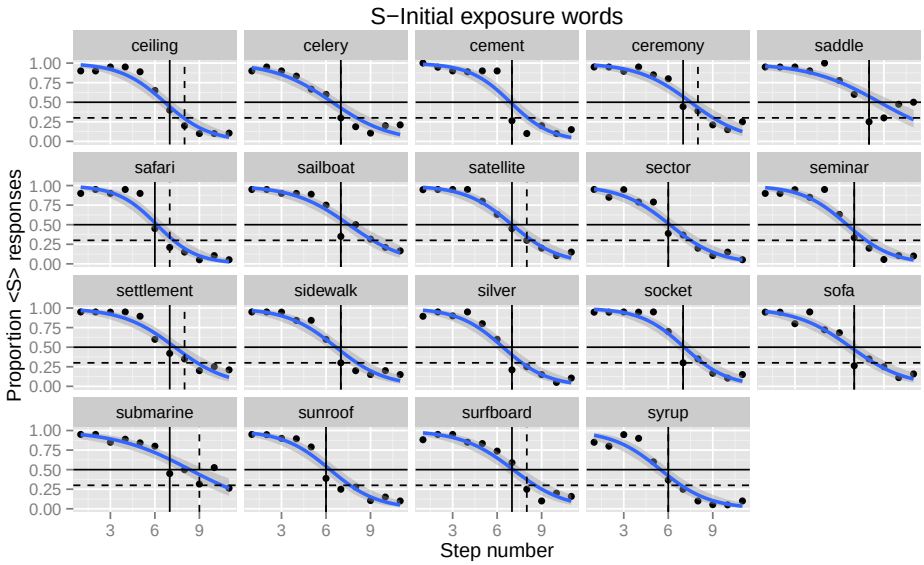

Figure 3.1: Proportion of word-responses for /s/-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model constructed from the responses.

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” 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 counter-

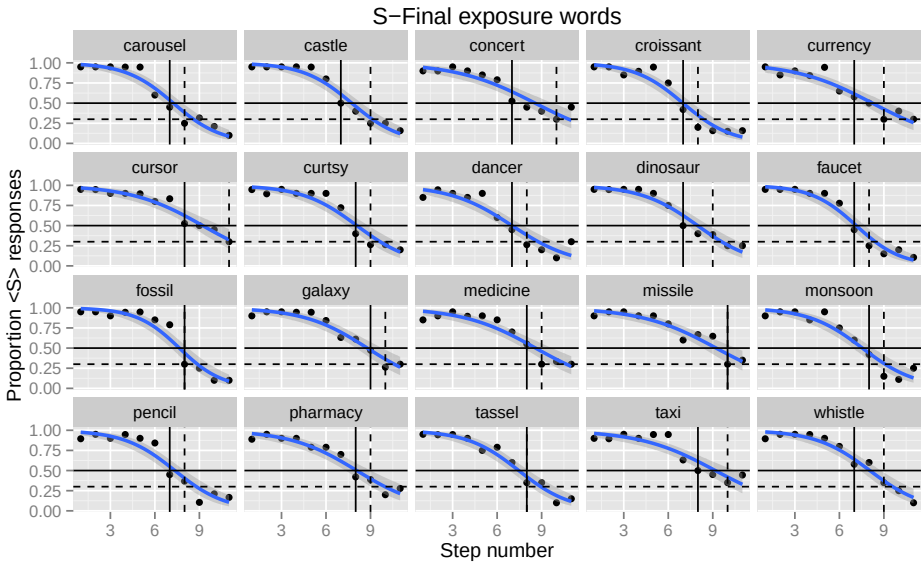

Figure 3.2: Proportion of word-responses for /s/-final exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model constructed from the responses

balanced 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. [27].

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.

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 syl-

lable, 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.

3.2.2 Results

3.2.3 Discussion

3.3 Experiment 2

3.3.1 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 ($SD = 0.9$).

3.3.2 Results

3.3.3 Discussion

Chapter 4

Cross-modal word identification

4.1 Motivation

4.1.1 Semantic predictability

The type of contextual predictability used in this experiment is known as semantic predictability. Sentences are highly predictable when they contain words prior to the final word that points almost definitively to the identity of the final word. For instance, the sentence fragment *The cow gave birth to the...* from Kalikow et al. [11] 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. Semantic predictability has been found to have effects on both production and perception, which will be discussed in the two following sections

Acoustic realization

In general, semantic predictability has a reductive effect on the acoustics of words. In Scarborough [29], 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 had lessened 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 [4] 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 [29], 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.

Intelligibility

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 [11]. 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 [19]. ¡¡I don't know if this is just due to differences in speaking rates/durations. Do you know of any studies (maybe based on sentences excised from sponataneous speech) where words would be said with similar durations or have contextual frames that normalize the duration differences for the listener?¿¿

4.1.2 Similarity to lexical bias

A key pillar of the motivation underlying this experiment is the notion that semantic predictability is similar to, yet distinct from, lexical bias/lexical predictability. Scarborough [29] found that lexical predictability and semantic predictability did not interact and had the same effect directions, but different effect sizes, with lexical predictability having a larger effect than semantic predictability.

One line of research has attempted to show whether the semantic predictability effects are part of perceptual and phonetic processing or if they result following

phonetic processing. Connine and Clifton [6] 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 consisten 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.

Two studies that have looked at semantic predictability in this paradigm are Connine [5] and Borsky et al. [2], and they found conflicting results. In Connine [5], the reaction time profile aligned more with the monetary bias in Connine and Clifton [6], with reaction time differences located at the end point and not the category boundary, but Borsky et al. [2] found the reverse with a profile more similar to lexical bias. ¡¡Examine differences between studies?¿¿

4.1.3 Attention?

I don't know of any work that has specifically looked at attention (to acoustics in particular) in the context of semantic predictability.

4.2 Methodology

4.2.1 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.

4.2.2 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 [11], 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 sh zh ch jh/). 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 /sh/. 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 [12]; 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. As in Experiments 1 and 2, the control and filler items were processed and resynthesized. 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.

4.2.3 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.

4.2.4 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.

4.3 Results

4.4 Discussion

Chapter 5

Conclusions

Stuff I have concluded

Bibliography

[1] 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, 8, 9

[2] 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 9, 21

[3] 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 14

[4] 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 20

[5] 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 21

[6] 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 21

[7] F. Eisner and J. M. McQueen. The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2):224–238, 2005. → pages 5

[8] 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, 4

[9] 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 9, 11

[10] A. Jesse and J. M. McQueen. Positional effects in the lexical retuning of speech perception, 2011. ISSN 1069-9384. → pages 8

[11] 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 9, 19, 20, 22

[12] 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 15, 22

[13] 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, 6

[14] T. Kraljic and A. G. Samuel. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2):262–268, Apr. 2006. ISSN 1069-9384. doi:10.3758/BF03193841. URL http://link.springer.com/10.3758/BF03193841. → pages 7

[15] 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 5

[16] 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 8

[17] 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 9

[18] P. Ladefoged and D. E. Broadbent. Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1):98–104, 1957. ISSN 00014966. URL
http://scitation.aip.org/content/asa/journal/jasa/29/1/10.1121/1.1908694. → pages 2

[19] 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 20

[20] S. L. Mattys and L. Wiget. Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160, 2011. → pages 12

[21] H. Mitterer, O. Scharenborg, and J. M. McQueen. Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2):356–361, 2013. → pages 7

[22] D. Norris, J. M. McQueen, and A. Cutler. Perceptual learning in speech, 2003. → pages 2, 3, 9

[23] 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 9, 12, 13

[24] 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 12

[25] 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 12

[26] 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 6

[27] 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 5, 15, 17

[28] 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 8

[29] 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 19, 20

[30] 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