

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

© Michael McAuliffe, 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

Preface	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Glossary	vii
Acknowledgments	viii
1 Introduction	1
2 Background	2
2.1 Perceptual learning	2
2.2 Rentention	2
2.3 Speaker dependency	3
2.4 Timecourse	3
2.5 Variation	3
2.6 Generalization	3
2.7 Dependencies	3
3 Lexical decision	4
3.1 Motivation	4
3.1.1 Lexical bias	4

3.2	Experiment 1	6
3.2.1	Methodology	6
3.2.2	Results	11
3.2.3	Discussion	11
3.3	Experiment 2	11
3.3.1	Methodology	11
3.3.2	Results	11
3.3.3	Discussion	11
4	Cross-modal word identification	12
4.1	Motivation	12
4.1.1	Semantic predictability	12
4.1.2	Similarity to lexical bias	13
4.1.3	Attention?	14
4.2	Methodology	14
4.2.1	Participants	14
4.2.2	Materials	15
4.2.3	Pretest	16
4.2.4	Procedure	17
4.3	Results	17
4.4	Discussion	17
5	Conclusions	18
	Bibliography	19

List of Tables

Table 3.1	Mean and standard deviations for frequencies and number of syllables of each item type	7
Table 3.2	Frequencies of words used in categorization continua	8

List of Figures

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.	9
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	10

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

2.1 Perceptual learning

A listener's perceptual system shifts rapidly; for instance, vowel perception can shift solely on the basis of the vowels in a preceding sentence [16]. Boundaries between fricative categories can be changed by exposure to relatively small amount of ambiguous fricative tokens, but they must be embedded within and associated with a lexical item in a language, and hearing ambiguous productions between two sound categories in nonwords does not induce perceptual learning [19]. However, besides the categorical distinction separating words from nonwords, little is known about the factors that influence linking an shifted or ambiguous production of a sound with a category through a lexical item. My dissertation investigates the factors which can contribute to associating ambiguous atypical productions with a specific sound category through lexical items.

Norris et al. [19]

2.2 Rentention

Kraljic and Samuel [11]

Eisner and McQueen [7]

2.3 Speaker dependency

Kraljic and Samuel [13]

Reinisch et al. [24]

2.4 Timecourse

?] [?]

Trude and Brown-Schmidt [27]

2.5 Variation

Sumner [26]

2.6 Generalization

Generalizing to other categories

Kraljic and Samuel [12]

?]

2.7 Dependencies

Context specificity

Kraljic et al. [14]

Kraljic et al. [15]

Reinisch et al. [23]

Clare [3]

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 [8]. 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 [22]. 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 [22] also found that lexical bias from monosyllabic words show greater response to experimental manipulations than trisyllabic words.

Pitt and Szostak [20] 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 [21], they found that, for monosyllabic word frames, token-final targets produce more stable results than token-initial targets.

Pitt and Szostak [20] 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 [18] 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 incorporated 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 [20].

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 tʃ dʒ/. 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 [2] 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 [10] 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 minimal 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 [24].

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

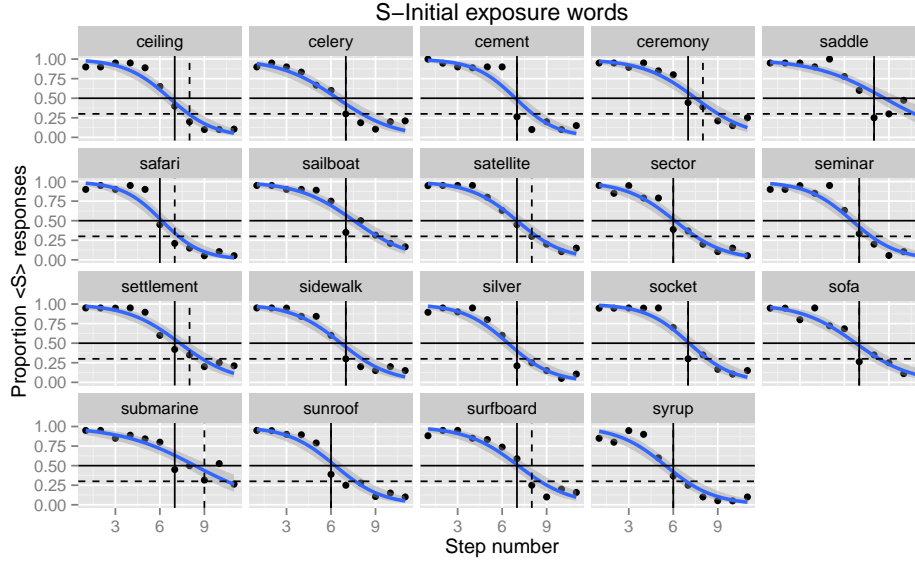


Figure 3.1: Proportion of word-responses for /s/-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line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

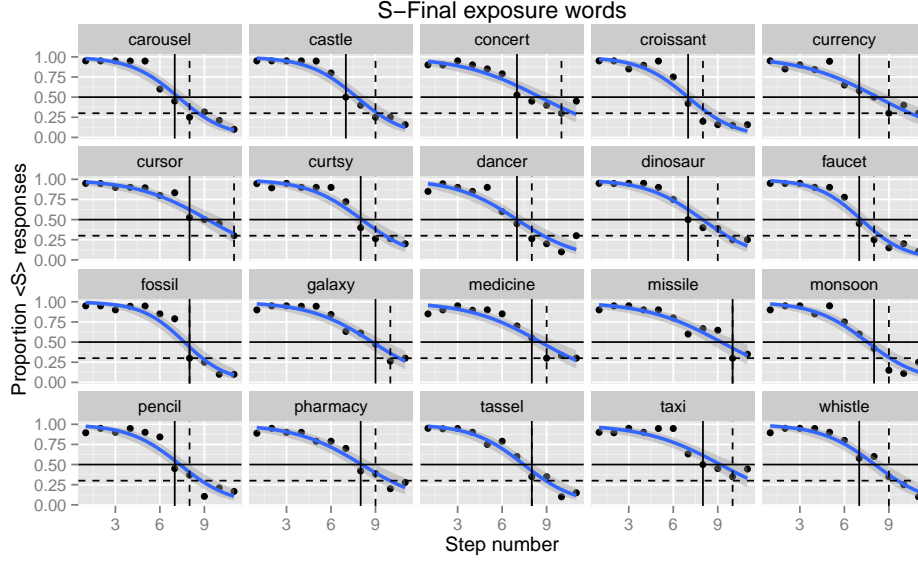


Figure 3.2: Proportion of word-responses for /s/-final exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line 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. [24].

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 be 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 critical 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. [9] 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 [25], 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 [25], showing temporal and spectral reduction in high predictability environments. However, speakers in the Midland dialect showed no such effect, and speakers of the Northern dialect showed more extreme Northern Cities shifting in 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 [9]. 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 [17]. 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 spontaneous 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 [25] 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 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.

Two studies that have looked at semantic predictability in this paradigm are Connine [5] and Borsky et al. [1], 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. [1] 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 sometimes 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 unpredictable of the final word. Unlike previous studies using sentence or semantic predictability [9], 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 [10]; 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 unpredictable 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] 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 14
- [2] 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 7
- [3] E. Clare. Applying phonological knowledge to phonetic accommodation, 2014. → pages 3
- [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 13
- [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 14
- [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 14
- [7] 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 2

- [8] 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 4
- [9] 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 12, 13, 15
- [10] 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 8, 15
- [11] 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 2
- [12] 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 3
- [13] 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 3
- [14] 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 3

- [15] 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 3
- [16] 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
- [17] 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 13
- [18] S. L. Mattys and L. Wiget. Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160, 2011. → pages 5
- [19] D. Norris, J. M. McQueen, and A. Cutler. Perceptual learning in speech, 2003. → pages 2
- [20] 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 5, 6
- [21] 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 5
- [22] 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 5
- [23] E. Reinisch, D. R. Wozny, H. Mitterer, and L. L. Holt. Phonetic category recalibration : What are the categories ? 2011. → pages 3
- [24] 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 3, 8, 10

- [25] 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 12, 13
- [26] 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 3
- [27] A. Trude and S. Brown-Schmidt. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27 (April 2013):37–41, 2012. URL <http://www.tandfonline.com/doi/abs/10.1080/01690965.2011.597153>. → pages 3