

# Connected Speech Processes and Lexical Access in Real-Time Comprehension

by

Mercedeh Mohaghegh

A thesis submitted in conformity with the requirements  
for the degree of Doctor of Philosophy

Department of Linguistics  
University of Toronto

© Copyright by Mercedeh Mohaghegh 2016

# Connected Speech Processes and Lexical Access in Real-Time Comprehension

Mercedeh Mohaghegh

Doctor of Philosophy

Department of Linguistics  
University of Toronto

2016

## Abstract

Connected speech can entail variability in the production of speech sounds. This can in turn create ambiguity at the lexical level. For example, the word *bean* in ‘*bean box*’ can sound like *beam* due to the phonological process of coronal-to-labial place assimilation. Previous studies have shown that listeners can perceptually compensate for place assimilation, but the extent of compensation, as well as the factors that guide this process, are a topic of ongoing debate.

This thesis explores listeners’ compensation for place assimilation from several novel perspectives. One specific concern involves potential differences across the sound classes affected by place assimilation (nasal versus oral stops), and whether these sound classes show similar patterns of compensation when the triggering phonological context (i.e., a word-initial labial consonant following the assimilated sound) is or is not present. A second issue is the extent to which effects of compensation are observed for these sound classes in the early moments of word recognition, and a third consideration is whether the English lexicon is structured in a way that minimizes confusion. An acoustic analysis, two word identification experiments, and two experiments using an eye tracking methodology combined with a priming paradigm are used to examine the production and perceptual processing of unassimilated and assimilated word-final

nasal and oral stops. A corpus analysis is also conducted to explore the structure of the lexicon with respect to the cases where place assimilation might result in lexical ambiguity.

I demonstrate that complete compensation for assimilation may be less likely to occur than previously assumed. However, the phonological context information clearly plays a role in compensation for assimilation even when the degree of assimilation is very strong. Further, the results of priming manipulations suggest that the precise nature of compensation may vary across sound classes. Finally, the structure of the lexicon seems to reflect the potential for confusion that results from coronal-to-labial assimilation in nasal and oral stops. Together, the findings suggest that, in addition to general auditory processing, inferential mechanisms and the statistical patterning of sounds within the lexicon play important roles in facilitating the recognition of assimilated words.

## Acknowledgments

Throughout the course of my studies as a PhD student, numerous people have contributed to my work and provided me with their support in various ways, to whom I owe a great deal of thanks. First, I would like to thank my supervisor, Craig Chambers, for always being there for me whenever I needed his guidance and support. His wisdom and knowledge combined with his patience and kindness helped me overcome the challenges and move forward in my project.

I am also grateful to my advisory committee members, Yoonjung Kang and Philip Monahan, who have helped me immensely with their valuable insight and practical advice throughout the project. I would also like to thank Elizabeth Johnson, Laura Colantoni and Benjamin Tucker for their useful feedback and encouraging comments, which made the last stage of finishing the thesis project a fruitful and enjoyable experience for me. A special thank you to Keren Rice for her kind support in various aspects of my academic and personal life, and for providing me with her wise words when I needed them the most. I would also like to thank Kathy Pichora-Fuller, Elan Dresher, Arsalan Kahnemuyipour, Alexei Kochetov, Daphna Heller, Diane Massam, Marshall Chasin, Jeffrey Steele and Michelle Troberg for all the encouragement, motivation and inspiration at different stages of my graduate studies that helped me with finding my path and following my dreams. And thanks to Mary Hsu for her support and friendship, for being my “big sister” when I was in need of one, and, of course, for being such an inspirational Zumba partner!

I have also been very lucky to be part of a wider community of wonderful individuals that spans across the departments of Linguistics, Language Studies (UTM), Psychology, and the PAL Lab (UTM). My friends, colleagues and fellow students have always provided me with their emotional and intellectual support throughout my journey as a graduate student. I would especially like to thank my good friends Maria Kyriakaki, Julia Yu-Ying Su, Yasaman Rafaat, Joanna Chociej, Dan Milway, Ross Godfrey, Matt Hunt Gardner, Christopher Harvey, Iryna Osadcha, Yu-Leng Lin, Manami Hirayama, Ross Krekoski, Ailis Cournane, Christopher Spahr, Alexandra Motut, Derek Denis, James Smith, Patrick Murphy, Youri Zabbal, Jessamyn Schertz, Aravind Kumar, Raheleh Saryazdi, Ewa Gambal, LeAnn Brown, Will Oxford, Richard Compton, and Sarah Clark who have touched my life and with whom I share wonderful memories.

A special thank you goes out to Payam Mehradnia, who stood by my side through the ups and downs of the graduate studies, especially supporting me emotionally and mentally during the final stages of the process.

And my parents, Farah Naghibzadeh-Jalali and Majid Mohaghegh. I will always be in debt to them for their endless love, devotion, and selfless acts that have supported me at every single stage of my life. Words cannot express how grateful I am to them!

**Thank you!**

## Table of Contents

<b>ACKNOWLEDGMENTS .....</b>	<b>IV</b>
<b>TABLE OF CONTENTS .....</b>	<b>VI</b>
<b>LIST OF TABLES .....</b>	<b>IX</b>
<b>LIST OF FIGURES .....</b>	<b>X</b>
<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
1.1 Spoken Word Recognition and the Issue of Variation.....	3
1.1.1 Phonological Variation: Place Assimilation .....	7
1.1.1.1 Place Assimilation in English: Nasal and Oral Stop Consonants .....	9
1.2 Theoretical Accounts of the Processing of Assimilated Sounds .....	13
1.2.1 Phonological Inference and Probabilistic Speech Perception.....	14
1.2.2 Feature Parsing.....	17
1.2.3 Perceptual Integration .....	20
1.3 Integration of the Contextual Information .....	22
1.4 Lexical statistics.....	25
1.5 Thesis structure .....	26
<b>CHAPTER 2 PREPARATION OF AUDITORY MATERIALS AND ACOUSTIC ANALYSIS .....</b>	<b>28</b>
2.1 Preparation of Auditory Materials .....	30
2.2 Acoustic Analysis .....	32
2.2.1 Formant Frequency .....	32
2.2.1.1 Nasal consonants.....	33
2.2.1.2 Stop Consonants.....	38
2.2.2 Consonant Duration .....	44
2.3 Discussion .....	47
<b>CHAPTER 3 FORCED CHOICE IDENTIFICATION STUDIES.....</b>	<b>50</b>
3.1 General Method and Materials.....	55
3.2 Experiment 1: Nasal Place Assimilation.....	60

3.2.1	Method .....	60
3.2.1.1	Participants.....	60
3.2.1.2	Materials .....	60
3.2.1.3	Procedure .....	60
3.2.2	Results.....	61
3.3	Experiment 2: Stop Place Assimilation .....	65
3.3.1	Method .....	65
3.3.1.1	Participants.....	65
3.3.1.2	Materials .....	65
3.3.1.3	Procedure .....	66
3.3.2	Results.....	66
3.4	Cross-Experiment Comparisons and General Discussion .....	69
<b>CHAPTER 4 VISUAL WORLD EYE TRACKING EXPERIMENTS.....</b>		<b>74</b>
4.1	General Method and Materials.....	76
4.2	Experiment 3: Nasal Place Assimilation.....	79
4.2.1	Method .....	79
4.2.1.1	Participants.....	79
4.2.1.2	Materials .....	79
4.2.1.3	Procedure .....	79
4.2.2	Results.....	80
4.3	Experiment 4: Stop Place Assimilation .....	90
4.3.1	Methodology .....	90
4.3.1.1	Participants.....	90
4.3.1.2	Materials .....	90
4.3.1.3	Procedure .....	91
4.3.2	Results.....	91
4.4	Cross-Experiment Comparisons and General Discussion .....	98
<b>CHAPTER 5 CORPUS STUDY: MINIMAL PAIRS ENDING IN /N/-/M/ OR /T/-/P/.....</b>		<b>105</b>
5.1	English Corpus Analysis.....	106
5.2	Discussion .....	109
<b>CHAPTER 6 CONCLUSIONS.....</b>		<b>111</b>

6.1	The Effect of Phonological Context on Perceptual Judgements.....	111
6.2	Nasal versus Oral Stop Place Assimilation.....	113
6.3	Structure of the Lexicon and Place Assimilation.....	119
6.4	Implications for the Accounts of Compensation for Assimilation .....	121
6.5	Concluding Remarks.....	125
<b>REFERENCES.....</b>		<b>128</b>
<b>APPENDICES .....</b>		<b>140</b>



## List of Tables

Table 2-1 Three measures of F2 frequency and the corresponding standard deviations (SD) for prime and target words ending in nasal consonants.....	35
Table 2-2 Three measures of F2 frequency and the corresponding standard deviations (SD) for prime and target words ending in stop consonants. ....	41
Table 2-3 Average combined duration of the word-final oral and nasal stops and the following labial stop and the corresponding standard deviation (SD) for prime and target words ending in stop consonants. ....	45
Table 3-1 Example prime and target stimuli from Experiment 1 in the original carrier sentences. ....	57
Table 5-1. Overall incidence, mean difference in lexical frequency, and mean semantic similarity index (LSA) for minimal pairs ending in /n/-/m/ or /t/-/p/ in English. ....	107

## List of Figures

Figure 2-1 Sample waveforms and corresponding spectrograms for <i>dine</i> [dajn] (upper panel), <i>dine</i> [dajn <sup>m</sup> ] (middle panel) and <i>time</i> [tajm] (lower panel).....	34
Figure 2-2. F2 frequency measures for word-final labial, coronal and assimilated nasals.....	37
Figure 2-3 Sample waveforms and corresponding spectrograms for <i>cat</i> [kæt] (upper panel), <i>cap</i> [kætp] (middle panel) and <i>tap</i> [tæp] (lower panel).....	39
Figure 2-4 F2 frequency measures for word-final labial, coronal and assimilated stops. ....	42
Figure 2-5 Average combined duration of the word-final stop consonants and the following labial stop for labial, assimilated coronal and unassimilated coronal oral stops (left panel) and nasal stops (right panel). ....	46
Figure 3-1. Example of visual display. ....	56
Figure 3-2. Percentage of correct responses for isolated target words ending in nasals (Exp. 1), across conditions. ....	62
Figure 3-3. Average response RT for isolated prime words ending in nasals (Exp. 1).....	63
Figure 3-4. Percentage of correct responses for isolated target words ending in oral stops (Exp. 2). ....	67
Figure 3-5. Average correct response RT for isolated prime words ending in oral stops (Exp. 2). ....	68
Figure 3-6. Percentage of correct responses for isolated target words ending in nasals (left panel, Exp.1) or oral stops (right panel, Exp. 2).....	70
Figure 3-7. Average correct response RT for isolated prime words ending in nasals (left panel) or stops (right panel). ....	72
Figure 4-1. Example of a critical display in which one word serves as the target, one as a competitor, one as a prime and two as distractors. ....	78

Figure 4-2. Percentage of correct responses for target words ending in nasal stops in isolation (left panel, Exp. 1) and in their carrier sentence (right panel, Exp. 3), across conditions. The error bars represent the standard error of the mean. .... 81

Figure 4-3. Average correct response RT for prime words in isolation (left panel, Exp. 1) or in their carrier sentence (right panel, Exp. 3) that ended in nasals. .... 83

Figure 4-4. Mean fixation proportion on buttons associated with the unassimilated targets ending in a nasal, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). .... 86

Figure 4-5. Mean fixation proportion on buttons associated with the assimilated targets ending in a nasal, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). .... 88

Figure 4-6. Percentage of correct responses for isolated target words (right panel, Exp. 2) and words in their carrier sentences (left panel, Exp. 4) ending in oral stops, across conditions. The error bars represent the standard error. .... 92

Figure 4-7. Average correct response RT for prime words ending in oral stops in isolation (left panel, Exp. 2) or in their carrier sentence (right panel, Exp. 4). .... 93

Figure 4-8. Mean fixation proportion to buttons associated with the unassimilated targets ending in an oral stop, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). .... 95

Figure 4-9. Mean fixation proportion on buttons associated with the assimilated targets ending in oral stops, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). .... 97

Figure 4-10. Percentage of correct responses for target words ending in nasals (left panel, Exp.3) or oral stops (right panel, Exp. 4) when presented in their carrier sentence. .... 101

Figure 4-11. Average correct response RT for prime words that end in nasals (left panel) or oral stops (right panel) when presented in their carrier sentence. .... 102

Figure 4-12. Average advantage of fixations to the target over the competitor words for words ending in nasals (left panel, Exp. 3) or ending in oral stops (right panel, Exp. 4), across conditions (within 200-600 ms after the onset of /b/ in *button*). ..... 103

## List of Appendices

Appendix A. Experimental stimuli used in critical trials in Experiments 1 and 3.....	140
Appendix B. Experimental stimuli used in critical trials in Experiments 2 and 4.....	141
Appendix C. Experimental stimuli used in filler trials in Experiments 1-4.....	142
Appendix D. Word frequency retrieved from SUBTLEXus database for the displayed prime, undisplayed prime competitor (if available), and displayed target and displayed target competitor words used in Experiments 1-4.....	144

# Chapter 1

## Introduction

Speech comprehension is natural, automatic and effort-free for most speakers when communicating in their native language(s). Yet the computational system that underlies successful language comprehension is extremely complex and involves the processing of information at various levels, from detailed fine-grained acoustic cues to pragmatic and contextual information. One essential part of this process is recognizing the words that comprise sentences. This involves the incremental mapping of the acoustic signal to lexical candidates as speech unfolds in time. However, the inherently variable nature of the speech signal poses major challenges for the recognition of spoken words (Connine & Pinnow, 2006; Jusczyk & Luce, 2002; Hale, Kisko, & Reiss, 2007; Monahan & Idsardi, 2010; Tucker & Warner, 2007). Random and regular changes in the pronunciation of individual sounds within and across word boundaries as a result of physiological, co-articulatory or language specific contextual effects can result in the blurring of contrasts among individual sounds. An even greater challenge arises in cases where alternative pronunciations of speech sounds can potentially result in lexical ambiguity. This dissertation focuses on the mechanisms involved in processing regular phonological variation in the speech signal and the factors that influence lexical access in the absence of clear contrasts at the level of speech sounds. Specifically, I focus on the influence of phonological context, characteristics of sounds, and the structure of the lexicon on the recognition of lexically ambiguous forms that have been affected by the phonological process of *place assimilation* in English.

Word-final place assimilation is a common phonological process in the world's languages (see section 1.1.1.1 for an articulatory description of the phenomena). Like other forms of phonological variation, this process involves a systematic mismatch between the speech signal and its abstract representation. This mismatch can be predicted based on the particular sound environment the segment is produced in. In English, for example, regressive place assimilation is known to affect coronal nasal and stop consonants /n, t, d/ when they appear in word-final (or syllable-final) position. These sounds are often described as taking the place of articulation of a following word-initial labial (or velar) segment (e.g., /b, g, m, ŋ/; Chomsky & Halle, 1968). For example, in 'cat box', the final /t/ in *cat* can assimilate to the place of articulation of the

following labial consonant /b/ and as a result, *cat* would be pronounced similar to the word *cap*. The process of place assimilation can result in various degrees of alteration in the surface form of the sound compared to its underlying form. Depending on the extent to which acoustic cues to the underlying coronal place of articulation remain, the results can range from relatively weak to strong or even complete assimilation (where the sound is indistinguishable from a true labial/velar). This poses a great challenge in processing connected speech, particularly when the alternative pronunciation also matches another existing word. For example, the word *cat* in '*cat box*' becomes perceptually similar to the word *cap* in '*cap box*' after undergoing place assimilation.

Previous studies have suggested different explanations for how the recognition system handles this type of phonological variation. One view, for example, has attempted to account for the perception of assimilated sounds by emphasizing the abstract representational features of sounds (e.g., Lahiri & Reetz, 2010). Within this framework, the role of the contextual information in compensating for assimilation is minimal, if at all relevant. The majority of previous studies, however, agree that phonological context plays an important role in the perception of (partially) assimilated sounds (e.g., Gaskell, 2003; Gow, 2001, 2002; Mitterer, Csepe & Blomert, 2006). Nevertheless, there is still controversy with regard to the recognition of assimilated forms that entail lexical ambiguity (especially when the assimilation is strong/complete) and whether the knowledge of one's language-specific phonological system assists in recovering the intended form and to what extent.

This dissertation addresses the following questions regarding the recognition of assimilated forms:

1. How does the recognition system make use of the available phonological context information alongside sub-lexical acoustic information when identifying an assimilated consonant? How is recognition affected when the contextual information is removed?
2. Do the inherent differences between oral and nasal stop consonants affect how listeners process assimilated forms? More specifically, does the recognition system treat sounds that are affected by a similar assimilatory process the same way even if they are from different natural classes such as nasal and oral stop consonants?

3. Is there any relationship between the structure of the lexicon and patterns of production and perception of place assimilation in different sound categories? If there is a relationship, do lexical statistics work against or in favor of resolving phonological variation?

In particular, the current study is organized around the concept of processing place assimilation by building on a body of research examining the role of phonological context and fine-grained acoustic cues in compensation for phonological variation. I examine a range of factors including the unique perceptual and acoustic characteristics of sound categories, the structure of the lexicon, and their relationships with the recognition of spoken words in real time. In the remaining sections of this chapter, I provide background for the core questions above and provide a review of the relevant literature on each topic. Stimulus preparation and an acoustic study on the experimental materials are described in Chapter 2. In Chapter 3, I describe two identification experiments focusing on the recognition of isolated words that end in either unassimilated or assimilated coronal oral or nasal stops. In Chapter 4, two experiments using an eye-movement priming paradigm are described. These experiments focus on the real-time recognition of the same unassimilated or assimilated words when presented in their original carrier sentences (i.e., where the triggering phonological context is provided). An analysis of English corpus data is discussed in Chapter 5, where the potential for words ending in assimilated nasal or oral stops to entail lexical ambiguity is compared. Finally, a general discussion of the findings of the research and final conclusions are provided in Chapter 6.

## 1.1 Spoken Word Recognition and the Issue of Variation

The comprehension of an utterance, an essential aspect of human communication, depends on a listener's ability to recognise and decode the smaller pieces that comprise it, namely words. In spoken language, the input is a stream of speech sounds produced by an individual speaker. The process of spoken word recognition acts as an interface between this continuous input and stored representations in the listener's mental lexicon (Gaskell & Marslen-Wilson, 1997, 1999; Lahiri & Reetz, 2002; Magnuson, Dixon, Tanenhaus & Aslin, 2007; McClelland & Elman, 1986; Norris, 1994). The incremental mapping of phonemic information to lexical candidates is understood to operate continuously and immediately, rather than being delayed until after the whole word is available to the listener (e.g., Warren & Marslen-Wilson, 1987). In addition, the process involves parallel activation of multiple lexical items. As successive sounds unfold, these



lexical items are considered to various degrees based on their match with the input stream and eventually one candidate will be chosen/win the ‘competition’. For example, by the time the partial string [kærəp] is heard, the signal would match with only one lexical candidate, namely the word *caterpillar*.

Traditionally, the input to spoken word recognition at the sub-lexical level was thought to have been provided by autonomous speech perception processes that operated at an earlier stage. These processes worked to map the acoustic signal onto more abstract phonetic and/or phonemic representations so that the resulting output could then be mapped onto lexical representations to access word forms. However, an ample amount of recent evidence, including studies using real time measures such as tracking eye movements, suggests an interactive relationship between the two levels of processing, whereby fine-grained acoustic cues are shown to affect higher levels of lexical processing (e.g., Dahan & Gaskell, 2007; Dahan, Magnuson, Tanenhaus & Hogan, 2001; McMurray, Clayards, Tanenhaus & Aslin, 2008; McMurray, Tanenhaus & Aslin, 2002; McQueen & Cutler, 2001; McClelland, Mirman & Holt, 2006) and, conversely, linguistic knowledge affecting lower level auditory parsing (e.g., Beddor, McGowan, Boland, Coetzee & Brasher, 2013; Dahan, Drucker & Scarborough, 2008; Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus & Aslin, 2003; McQueen & Cutler, 2003; Viswanathan, Fowler & Magnuson, 2009).

The processes involved in spoken word recognition are amazingly robust, making it seem like an effortless task despite the fact that everyday language comprehension most often happens under potentially challenging conditions. Among the known factors that result in processing difficulties is the lack of invariance of the input speech signal (e.g., Jusczyk & Luce, 2002; Weber & Scharenborg, 2012). As mentioned earlier, to comprehend spoken words, the speech input is matched against the stored lexical candidates; however, the specific realisation of a given speech sound is, for most part, situation-specific and can dramatically change due to both idiosyncratic factors, such as speech style and accent, rate of speech or even physiological characteristics of speakers’ vocal tracts (e.g., female vs. male), and more predictable and systematic factors such as co-articulation or phonological context. Variation in the acoustic signal potentially poses a significant problem for mapping the input onto abstract representations and ultimately, lexical recognition. Still, listeners perceptually overcome and even adapt to these variations for successful communication in everyday language use (e.g., Norris, McQueen & Cutler, 2003).

An example of a challenge listeners face arises from studies on talker-specificity that shows that efficiency (e.g., higher accuracy level) in speech perception tends to be lower when there is talker variability: a quite common situation in natural speech communication. Listeners perceive talkers' speech even if they have different accents under normal or degraded ambient conditions (e.g., with background noise) significantly more accurately if they were familiar with the speaker (Magnuson & Nusbaum, 2007; Mitterer & Reinisch, 2013; Nygaard & Pisoni, 1998; Trude & Brown-Schmidt, 2012). Several studies also report lower identification scores in conditions where speech rate varied from trial to trial (Norris, McQueen & Cutler, 2003; Sommers, Nygaard & Pisoni, 1994). Such effects of variation become especially relevant when they result in a change in an acoustic cue associated with a speech sound. For example, voice onset time (VOT), which is a primary cue to the voicing distinction in stops, can change based on speech rate (Toscano & McMurray, 2012). Shorter VOT is associated with voiced/unaspirated stop consonants, while longer VOT with voiceless/aspirated stops; however, as a temporal cue, VOT duration might decrease in fast speech, which, in the absence of other cues (such as vowel length), can potentially result in perceptual ambiguity or even lexical ambiguity (e.g., *bit* vs. *pit*).

Other evidence comes from studies examining variation due to influence of phonetic context. A distinction is sometimes made between variation resulting from rather mechanical (and unintentional) effects of co-articulation that create graded differences in the character of speech sounds, and the language-specific regular type of variation due to phonological processes that can potentially result in a categorical neutralization (Halle & Adda-Decker, 2011; Ohala, 1993). The mechanics of speech production require movement of the articulators from one position to another within the vocal tract. In casual speech, such rapid movements often result in overlap of the articulators, and can potentially result in a shift in the category boundary of a sound. For example, in production of [k] in [kæɹəˌpɪlə] 'caterpillar', the tongue dorsum is more anterior in the oral cavity than when producing [k] in [ku] 'cool', in which its position is more posterior. This is because in the former case [k] precedes the front vowel [æ] but precedes a back vowel [u] in the latter case. Effects of co-articulation are most often found in adjacent sounds (Beddor, McGowan, Boland & Coetzee, 2013; Dahan, Magnuson, Tanenhaus & Hogan, 2001; Warren & Marslen-Wilson, 1987), although long-range effects (e.g., across syllable or word boundary) have also been reported. For example, vowel-to-vowel co-articulation effects have been shown across word boundaries where the acoustic properties of a vowel (e.g., [o] in *pole*) is linked to

the properties of a preceding vowel produced within the same phrase ([ʌ] in “pick up a pole”; Coleman, 2002; Tobin, Cho, Jenet & Magnuson, 2010). As noted above, co-articulatory effects are usually characterized as gradient and incomplete. Phonological processes on the other hand, have been traditionally described as resulting in complete neutralization of the sound categories that are involved. For example, in colloquial English, ‘got you’ /gɑt ju/ is pronounced as ‘gotcha’ [gɑtʃɑ] (/t/ → [tʃ]) as a result of full palatalization, a common assimilatory process that changes the place of articulation of a consonant when it is adjacent to a palatal glide (or a high-front vowel) to a more central region of the vocal tract (Bateman, 2011; Nolan, 1992).

According to the above distinction, unlike the process of overcoming co-articulatory effects, which is believed to be related to the general auditory system and therefore not tied to linguistic processing *per se*, phonological variation is often understood to be processed at a more abstract level, based on listeners’ language-specific knowledge (see Poeppel & Monahan, 2011, for a review). This is because, even though co-articulatory processes can differ in form from language to language, their effects, which are tied to the mechanics and the limits on the movements of the articulators, are often considered to be language universal. On the other hand, phonological processes, such as flapping, consonant deletion, vowel nasalization, place assimilation, and epenthesis, show a stronger tendency to differ from language to language both in their occurrence and in degree (Goldsmith, Riggle & Alan, 2011; Chomsky & Halle, 1968; Ohala, 1993). For example, the full palatalization mentioned earlier is a common phonological process found in many languages, including English. However, the same pattern is not attested in languages such as Turkish or Navajo, indicating its language-specific nature (Bateman, 2011).

The distinction between co-articulatory versus phonological variation is not without controversy. Several recent studies have shown that one traditionally assumed characteristic of phonological variation, namely a complete and categorical change of the target sounds, might be less common than originally assumed. Studies on the perception and production of sounds in connected speech have found that, in many cases, phonological processes result in an incomplete change where the surface form of the target sound retains residual cues that can potentially reveal its underlying identity (Gow, 2001 and 2002; Dilly & Pitt, 2007; Dinnsen & Charles-Luce, 1984). Another distinction between phonological versus co-articulatory variation that has been called into question is the language specificity of the phonological processes. A number of studies in recent years on the perception of variation caused by phonological processes have shown little to no

effect of language background at the perceptual level (e.g., Mitterer, Csepe & Blomert, 2006; Gow, 2003). However, other studies have found opposite effects, showing that listeners from different language backgrounds tend to have a higher perceptual accuracy when encountering variation that are specific to their language (e.g., Clayards, Niebuhr & Gaskell, 2015; Darcy, Peperkamp & Dupoux, 2007).

### 1.1.1 Phonological Variation: Place Assimilation

The focus of the current study is on a particular type of regular speech variation, namely place assimilation. Assimilatory processes, and place assimilation in particular, are quite common in the world's languages. Assimilation in connected speech affects adjacent as well as distant speech segments, causing them to become more similar to each other in one or more of their articulatory or acoustic features such as voicing, place or manner (Clayards et al., 2015; Ernestus, 2003; Kuzla, Cho & Ernestus, 2007; Weber, 2001). Assimilation as such can be progressive, as when a trigger segment affects a sound that follows it (e.g., /z/ → [s] in 'rats' /ɪæʔ+/z/ pronounced as [ɪæʔs]), or regressive, as when the triggering sound follows the target segment as in place assimilation in English (e.g., /t/ → [t<sup>k</sup>]<sup>1</sup> when 'cute girl' is pronounced as [kjut<sup>k</sup>gɪl]).

Traditionally, phonological theories assumed place assimilation in English to be a phonological process that results in categorical and complete neutralization (e.g., Chomsky & Halle, 1968). However, as was mentioned in the previous section, recent research has provided empirical evidence indicating that assimilated sounds, such as a /t/ undergoing coronal-to-labial place assimilation, can carry various degrees of acoustic and articulatory properties of both the canonical (i.e., alveolar), as well as surface labial or velar places of articulation (Browman & Goldstein, 1990; Dille & Pitt, 2007; Gow, 2001; Gow, 2003). In Gow (2001, 2003), acoustic measurements of formant transition cues from the preceding vowel to an assimilated nasal recorded from phonetically naïve speakers showed that formant frequency transitions, especially F2, fall somewhere between measures associated with underlying coronals and those for underlying labial or velar nasals. Studies on the articulation of assimilated versus unassimilated

---

<sup>1</sup> The superscript notation has been used in this paper (e.g., [t<sup>k</sup>] or [n<sup>m</sup>]) to indicate partial (rather than complete) coronal place assimilation.

sounds using X-ray, EMMA or EPG techniques also support the idea that such modifications are mainly gradient in nature due to articulatory gestures overlapping in time (Browman & Goldstein, 1989, 1990; Tiede, Perkell, Zandipour & Matthies, 2001). On the other hand, in their corpus study of spoken English (Buckeye Corpus of Conversational Speech), Dilley and Pitt (2007) measured transition cues from the preceding vowel to the following assimilated consonant and reported that complete place assimilation does in fact occur in some cases. However, the possibility of an assimilation being incomplete can never be ruled out altogether merely based on articulatory and acoustic measures, as there is always a chance that a relevant acoustic cue to the consonant's underlying place of articulation has not yet been taken into consideration.

The presence of residual acoustic cues from incomplete place assimilation has important implications for perceptual accounts of the mechanisms involved in dealing with this variation. This is because these acoustic cues can potentially be used in recovering the underlying alveolar place of articulation of assimilated consonants. In fact, Gow (2002) found an effect of residual acoustic cues even in the perception of strongly assimilated forms. He conducted a series of experiments using a cross-modal form priming paradigm. In these experiments, the listeners would hear a prime word in a sentence (e.g., ...*right berries*...) and, 100 ms after the offset of the prime word (i.e., *right*), would be presented with lexical decision probes that were either phonologically identical or similar (e.g., RIGHT or RIPE), or unrelated (HEN or HEM) to the prime. The participants' task was to make a lexical decision upon hearing the probe item. Gow examined the priming effects and showed that assimilated words that can potentially be lexically ambiguous (e.g., assimilated *right* that sounds like *ripe*) in the relevant triggering context (e.g., *right berries*) only primed identification of words ending with a coronal sound (i.e., the assimilated *right* primes *right*, but not *ripe*) even when the stimuli used as prime words were rated as strongly assimilated by naïve listeners (i.e., sounded similar to *ripe*). On the other hand, unassimilated words that ended with a non-coronal sound in a context triggering assimilation (e.g., *ripe berries*) primed only words ending in a non-coronal consonant (i.e., *ripe*). Consequently, Gow (2002) argued that the presence of residual acoustic cues to the underlying coronal place in the assimilated sounds (and the lack of such cues in true labial sounds) explains the bias observed in the priming effect. Gaskell and Snoeren (2008), on the other hand, provided evidence for compensation even when assimilation was perceptually judged by listeners as

complete (i.e., when no traces of acoustic cues to the underlying place of articulation of the assimilated sound could be perceptually identified). Using a wide range of speakers to naturally produce assimilated and unassimilated words in viable and unviable contexts for assimilation, they found that complete assimilation does occur in natural speech and when it occurs, the viability of the following phonological context for assimilation can influence listeners' perception of the sounds. For example, listeners showed a higher tendency to (mis)identify *rum*, produced with a canonical word-final labial consonant, as *run* in “a quick rum picks you up”, where the phonological context is viable for assimilation, but not when the context is unviable (e.g., “a quick rum does you good”). This was especially true if the semantic context also favored the word ending in the coronal consonant. The viability effect observed in such cases cannot be explained based on the surface acoustic cues to place of articulation, as the cues should favor the adoption of a labial (or velar) interpretation. Therefore, Gaskell and Snoeren interpreted such perceptual tendencies to be a result of probabilistic learning and due to listeners' familiarity with the sound alterations that are most common in their native language, such as final coronals sounding similar to the following labial or velar consonants in English.

In the following section, I review the process of English place assimilation for two groups of sounds, nasal and stop consonants. Acoustic and typological differences between the two groups of sounds suggest that place assimilation in oral and nasal stops might not be exactly the same process. However, in most theoretical accounts of compensation for place assimilation, no distinction has been made between place assimilation in nasals versus oral stops.

### 1.1.1.1 Place Assimilation in English: Nasal and Oral Stop Consonants

Nasal place assimilation in English is a regressive process that results in an underlyingly coronal nasal sound being produced at or near the same place of articulation as that of a following labial or velar consonant. In the production of coronal nasals, the tip of the tongue approaches the alveolar ridge and completely blocks air passage through the oral cavity while the nasal cavity remains open. In rapid connected speech, however, the tongue either completely or partially skips movement toward the alveolar ridge and the articulators move to form the stricture for the next sound while the nasal cavity is still open. The amount of overlap in articulation in turn results in different degrees of place assimilation. This process can affect sounds both across morphemes and word boundaries. For example, the addition of the prefix *in-* ending in the

coronal nasal /n/ to the root morpheme *possible*, results in /n/ taking the labial place feature and the word being pronounced as [ɪm-pəsɪbəl]. Another example of this process is when a word like *nine* /najn/ precedes a word such as *beans* [binz] or *girls* [gəɪlz] in connected speech and is pronounced as [najn<sup>m</sup>binz] or [najn<sup>n</sup>gəɪlz] respectively. Here, the word-final coronal nasal assimilates to the place of articulation of the following labial or velar consonant across a word boundary.

Both regressive and progressive nasal place assimilation are common across languages. In some languages, the process is mandatory. In Japanese, for example, the place of articulation of a nasal always needs to match with the place of articulation of the following consonant resulting in obligatory regressive place assimilation (e.g., *tombo* ‘dragonfly’ vs. *kondo* ‘this time’; Weber, 2002). In English as well as some other languages (e.g., Dutch), on the other hand, nasal place assimilation is an optional process. Although nasal place assimilation is a well-known phonological process in English, previous studies on the rate of production of assimilated consonants in casual connected speech have reported that it in fact happens relatively infrequently. For example, in their study of a spontaneous speech corpus, Dilley and Pitt (2007) reported only 20% of the final alveolar nasals in word-final position to be assimilated when 73% of the nasals were not altered and the remaining 7% were labeled as omitted. Factors such as speech style (e.g., casual vs. careful), speaking rate or even social factors, such as gender or social class, can affect the rate of production of assimilated forms (e.g., Li & Kaiser, 2012; Nguyen, 2008).

Place assimilation in oral stops is also very common cross-linguistically. The process is similar to place assimilation in nasals in that the canonical place of articulation of a stop consonant changes and becomes perceptually and acoustically more similar to the place of articulation of the following consonant. In English, the process results in a final coronal stop assimilating to the following labial or velar consonant the same as in nasal place assimilation (e.g., *wet paint* pronounced as [wet<sup>p</sup> peɪnt]). The overlapping gestures resulting from co-articulation of the stop with the following consonant are also mechanically similar to nasal place assimilation except that the nasal cavity is closed during the production of oral stops. Since the oral cavity is also closed during the closure portion of oral stops, no acoustic information is available from the onset of the closure to the release burst other than voicing and closure duration. For nasals however, formants and anti-formants are present for the duration of the closure due to air flowing

through the nasal cavity. This means that some acoustic information regarding the place of articulation of the nasal might still be present during closure.

In terms of acoustic properties as well as cross-linguistic typology, the process of place assimilation is also reported to be different for nasal and oral stops. Typological studies show that, relative to nasals, oral stops assimilate much less frequently (Jun, 1995; Hura et al. 1992; Nolan & Kerswill, 1990; Dilley & Pitt, 2007; Zimmerer, Reetz & Lahiri, 2009). Winters (2002) indicated that in all languages that have been studied so far, there are no languages in which stop place assimilation is found but nasal place assimilation is not found. On the other hand, it is quite possible to find languages where only nasals assimilate and stops do not (e.g., Chiyao, Ponapean, Yoruba) or languages where both stop and nasal place assimilation are common (e.g., Malayalam, Catalan, Korean). Even in languages with both groups of sounds taking part in assimilation, stops are often reported to be assimilated less frequently than nasals in natural speech. For example, work by Zimmerer et al. (2009) investigating the Kiel speech corpus of German showed that, across lexical and function words, the majority of assimilated cases belonged to nasals (89%) while stops (7%) and fricatives (3%) assimilated far less frequently. In English, Dilley and Pitt (2007) also reported finding assimilated oral stop consonants in only 7% of the cases where place assimilation context was available. However, oral stops often did not appear in their canonical form either: Voiced coronal stops appeared in their canonical form in 51% of the tokens and, interestingly, voiceless coronal stops were unaltered in only 25% of the cases where assimilation context was provided and were produced as glottal stops or were completely deleted 68% of the times. This indicates that even though, in statistical terms, English listeners are more often exposed to assimilated nasals compared to assimilated oral stops, they encounter more coronal stops in their various reduced surface forms (deleted, glottalized or assimilated) than nasals when these sounds occur in an assimilation-triggering context (i.e., preceding a labial or velar consonant). This in turn might potentially influence listeners' expectations and perceptual judgement when they hear an ambiguous stop or nasal consonant.

One possible reason behind the asymmetry in place assimilation in nasals versus oral stops is the difference in the acoustic cues to their place of articulation. Previous studies have suggested that formant transition cues from the preceding vowel that are associated with the place of articulation of the consonant are generally perceptually less salient for nasals compared to stops. This is because the vowel usually becomes nasalized in the former group and this nasalization in



turn obscures the transition formants (Nolan and Kerswill, 1992; Kawahara & Garvey, 2014; Ohala & Ohala, 1993; Mohr & Wang, 1968). From a perceptual standpoint, the lower saliency of the acoustic cues to place of articulation of nasals generally renders those cues as less valuable. Consequently, place assimilation becomes less “costly” and more common for nasals than for consonants such as oral stops, where changes to the acoustics of place of articulation are more perceptible. The weak perceptibility of nasal place of articulation compared to stops has been shown in studies of speech perception (Gaskell, 1994; Gaskell & Snoeren, 2008; Jun, 2004; Kawahara & Garvey, 2014). For example, in a series of perceptual similarity judgement and identification experiments, Kawahara and Garvey (2014) showed that place of articulation of nasals in coda position are perceptually much more similar compared to the place of articulation of either released or unreleased oral stops and proposed a hierarchy for the perceptibility of place contrast as follows: voiceless stops > voiced stops > nasals. In the two similarity judgment tasks, listeners were presented with pairs of sounds (two nasals or two released or unreleased stops) in the coda position that minimally differed in terms of their place of articulation (labial, coronal or dorsal) and were asked to judge the pair for their similarity (e.g., [am-an] or [at-ak]). The results supported the proposed perceptibility hierarchy regardless of whether the stops were clearly released or not. In the identification tasks, the same set of stimuli was presented to the participants with the addition of background cocktail party noise to mimic realistic listening conditions. On each trial, the participant’s task was to choose one of the two visual options on the computer screen (e.g., “am” or “an”) according to which best matched the auditory stimulus they heard. The proposed perceptibility hierarchy was observed again, confirming the weaker perceptual saliency of place of articulation in nasals compared to stops, especially voiceless stops. This pattern of results was maintained even when the stimuli were presented in a pre-consonantal position (e.g., *am* heard in *amga*), although identification scores were overall lower compared to the no-context condition suggesting that a perceptual masking has occurred when the coda stop consonants were followed by another consonant.

The effect of the following phonological context that triggers place assimilation on the recognition of assimilated words has been the focus of a number of studies. Previous research has shown in fact that such contextual information can benefit the recognition of lexical items that end in assimilated consonants (e.g., Bien & Zwitserlood, 2013; Dilley & Pitt, 2007; Gaskell & Marslen-Wilson, 2001; Gow, 2001 & 2002; Mitterer, Csepe, Honbolygo & Blomert, 2006;

Snoeren, Segui & Halle, 2008; Weber, 2001). Results of offline identification tasks as well as online tasks such as gating, priming or eye tracking have shown that when the following context is viable for assimilation to occur (in English, either a labial or a velar consonant), listeners are faster and more accurate at recognizing the assimilated lexical item than when the following context is not viable (i.e., when it is a coronal consonant; e.g., Gaskell & Marslen-Wilson, 1998; Gow & McMurray, 2007). For example, in a series of studies using stimuli that contained both coronal nasal stops (e.g., *green bag*) and coronal oral stops (e.g., 'cat box'), Gow and colleagues reported that providing a viable phonological context, as opposed to an unviable context, for assimilation facilitates the recognition of assimilated words. In addition to the regressive effect, they also showed that assimilated sounds have a progressive effect, leading the listener to anticipate a consonant with a particular place of articulation (velar or labial) to follow the assimilated consonant (Gow, 2001, 2002, 2003; Gow & Im, 2004; Gow & McMurray, 2007).

Based on these and similar other studies, several theoretical accounts have been proposed concerning the processing mechanisms involved in the perception of assimilated sounds. In the following section, these accounts are introduced and discussed in more detail.

## 1.2 Theoretical Accounts of the Processing of Assimilated Sounds

Over the last decade, a number of studies have examined the mechanisms involved in the perception and recognition of assimilated words (Clayards, Niebuhr & Gaskell, 2015; Gaskell & Marslen-Wilson, 1998; Gow, 2001, 2002, 2003; Gow & McMurray, 2007; Lahiri & Marslen-Wilson, 1991; Lee & Pater, 2010; Mitterer & McQueen, 2009; Zimmerer, Reetz & Lahiri, 2009). The processing accounts based on this work range from being purely representational (e.g., Featurally Underspecified Lexicon model; Lahiri & Reetz, 2002) to those that place an emphasis on general perceptual mechanisms (e.g., perceptual integration; Mitterer, Csepe & Blomert, 2006). For example, the Featurally Underspecified Lexicon (FUL) model (Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002, 2010) proposes that the underspecification of lexical representations for certain phonological features allows recognition processing mechanisms to accommodate phonological variation, such as vowel nasalization or place assimilation. Based on the FUL model, in a language like English, the place of articulation of a segment can be specified for only either the [labial] or [dorsal] feature. However, coronal segments, such as alveolar consonants, are not specified for a [coronal] feature and as a result would be susceptible

to taking the [labial] or [dorsal] feature of the surrounding sounds on the surface, as is the case in assimilation. An important claim of an underspecification account is that phonological context has no direct role in recognition of assimilated forms: Any surface variation is going to be associated with a relevant underspecified feature (e.g., [coronal] for place of articulation) hence solving the recognition problem within (rather than across) the word boundary (i.e., lexical representation). For example, based on the FUL model, the input [stip], which ends in a sound with the specified feature [labial], would “match” and activate the lexical representation of *steep* but an input such as [stit] would “mismatch” the same lexical representation. On the other hand, [spit] and [spip] “do not mismatch” the lexical representation of *spit*, which ends in a sound with the underspecified feature [coronal], suggesting they can both potentially activate *spit* regardless of the following phonological context.

However, the majority of studies on the processing mechanisms involved in dealing with place assimilation agree that listeners use information regarding the viability of the phonological context to compensate for this type of variation, where relevant contextual information occurs across word boundary. Still, there remains a lot of controversy about the nature of this process, the precise mechanisms involved in dealing with phonological variation (and the ambiguities caused by it), and whether the same mechanisms apply to related phonological processes. In this section, I discuss the three major accounts proposed for understanding how listeners manage the problem of place assimilation, namely the phonological inference account (Gaskell, 2003; Gaskell & Marslen-Wilson, 1998, 2001), the feature parsing account (Gow, 2002, 2003) and the perceptual integration account (Mitterer, Csepe & Blomert, 2006).

### 1.2.1 Phonological Inference and Probabilistic Speech Perception

The phonological inference model for assimilation was put forward by Gaskell and colleagues (Gaskell, 2003; Gaskell & Marslen-Wilson, 1995, 1998, 2001; Gaskell & Snoeren, 2008). According to this account, there is a relationship between an assimilated consonant and the following phonological context that triggers the assimilation, and this relationship is modulated by listeners’ language experience. Specifically, the recognition system learns the statistical probability of a phonological alteration that takes place in a given language and uses that probabilistic knowledge to compensate for this alteration. In English nasal place assimilation for example, listeners frequently encounter a scenario in which a word-final coronal nasal would

sound like a labial nasal when it is followed by a word-initial labial consonant (e.g., *green* sounds like *greem* in “greenu bag”) and not when it is followed by a coronal consonant (e.g., *green tag*), and therefore they use that probabilistic information in “undoing” the effects of assimilation.

In a study by Gaskell and Marslen-Wilson (1996), the effect of having a phonological context that licenses place assimilation as opposed to one that does not (the “viability effect”), was demonstrated using a cross-modal repetition priming task. The authors auditorily presented unassimilated (e.g., *lean* [lin]) and completely assimilated (e.g., [lim]) forms of prime words in viable (e.g., *lean bacon*) versus unviable (e.g., *lean gammon*) phonological contexts for coronal-labial place assimilation within sentences. The visual target, which was always a printed word corresponding to the underlying form of the word (e.g., *lean*), was presented at the offset of the prime. Participants were then required to provide a lexical decision response to the visual target. The authors observed a facilitatory priming effect, as evidenced by faster lexical decisions for the visual target LEAN in the assimilated-viable context condition, whereas the assimilated-unviable context did not render the same priming effect. The regressive phonological inference account proposes that, since the strong assimilation results in the place of articulation of the final consonant changing from coronal to labial, auditory information alone cannot explain the observed patterns of priming effects. Instead the sound is first perceived based on the surface acoustic cues (i.e., as a labial) and then probabilistic knowledge modifies this perception at the phonological representation level to access the coronal feature based on the listeners’ experience that underlyingly-coronal nasals and stops might sound labial if they appear before a labial consonant. Because a stronger priming effect was observed for assimilated words compared to assimilated nonwords (e.g., *rean<sup>m</sup>*), this is understood as a top-down effect of using knowledge of the lexicon in processing auditory input.

Gaskell (2003) adds an auxiliary dimension to his probabilistic connectionist model of phonological inference to incorporate listeners’ sensitivity to acoustic details in the case of graded assimilation, which, as noted earlier, is the most commonly-found form of assimilation. On this modified account, weakly assimilated forms might carry some acoustic cues to their underlying place of articulation as well as some cues to the place of articulation of the following sound. This can explain the anticipatory effect for the identity of the following phonological context, as was discussed earlier in Section 1.1.1.1 (e.g., Gow 2001, 2002). The regressive

viability effect, whereby the viability of the following phonological context helps with identification of place of articulation of the assimilated consonant, is argued to be relatively weaker when the target consonant is weakly assimilated. However, weakly assimilated coda consonants show a stronger anticipatory effect by helping with identification of the place of articulation of the following consonant that triggers assimilation. On the other hand, the more strongly assimilated forms are recognised through regressive phonological inference and a probabilistic interpretation. For extreme cases of strong assimilation (complete assimilation), however, Gaskell proposed a decrease in the influence of the following phonological context.

Place assimilation can be especially problematic when it results in lexical ambiguity (e.g., *run* sounding similar to *rum*), particularly in cases where assimilated sounds undergo a categorical change. While in the recognition of assimilated words such as *lean* in “leanu bacon”, top-down lexical biases might help identify [lim] as the word *lean* rather than the non-word *lean*, such biases cannot help identify the underlying form of assimilated words that result in lexical ambiguity, such as *rum* in “run picks”, where both *run* and *rum* are real words. In such cases, phonological context information has been found to be insufficient to recover the underlying form, and therefore other types of information would be necessary for disambiguation to occur. Gaskell and Marslen-Wilson (2001) conducted a cross-modal priming task and used stimuli that had the potential for ambiguity after undergoing place assimilation. This time the authors compared neutral versus biased sentential contexts. They found that when the assimilation is complete and the sentential context is neutral (e.g., “I think a quick *rum* picks you up”), listeners tend to select the surface form of the lexical item as the intended form regardless of the viability of the following phonological context and therefore priming was only found for the labial item (e.g., *rum*) but not the coronal form (e.g., *run*). However, when the sentential context favoured the coronal interpretation over the labial one (e.g., “It’s best to start the day with a burst of activity; I think a quick *rum* picks you up”), they observed priming for both forms in a viable phonological context (e.g., *rum picks*) and only for the labial form in an unviable phonological context (e.g., *rum does*). These results demonstrate how higher-level sources of information, such as sentential context, might be necessary when the phonological context alone does not provide sufficient information to disambiguate the assimilated lexical item.

A crucial prediction of the phonological inference account is that compensation for assimilation is language-dependent and listeners can overcome variations that are common in their native

language better than variations that do not occur in their language. This view is supported by the results of studies on the perception of assimilated sounds that show listeners with different language backgrounds treat acoustically similar assimilated forms differently (e.g., Clayards, Niebuhr & Gaskell, 2015; Darcy, Ramus, Christophe, Kinzler & Dupoux, 2009; Mitterer, Kim & Cho, 2013). For example, Darcy et al. (2009), tested French and English listeners' perception of two phonological processes of French voicing assimilation and English place assimilation (the first is not found in English and the second is not found in French). They conducted two word detection tasks using stimuli in the native languages of the two groups of listeners. The production of the stimuli was deliberately modified to include voicing assimilation in English sentences and place assimilation in French sentences. Their results indicated that listeners are in fact better at compensating for the type of assimilation that occurs in their native language rather than the one that does not. Other studies have however reported some degree of language-independent effect in addition to language-dependent effects in compensation for various types of assimilation, suggesting that both phonological as well as perceptual mechanisms might be involved in coping with such variations (Clayards, Niebuhr & Gaskell, 2015; Mitterer, Csepe & Blomert, 2006).

### 1.2.2 Feature Parsing

The feature parsing account, proposed by Gow and colleagues, is based on the assumption that assimilation is a result of co-articulation and mechanisms involved in recovering assimilated sounds are related to the general auditory system (e.g., Gow, 2001, 2002; Gow & Im, 2004; Gow & McMurray, 2007; Gow & Segawa, 2009). As noted earlier, studies by Gow and colleagues appear to assume that there will always be some residual cues to the underlying place of articulation in assimilated consonants. Thus, place assimilation is a graded process and the output of the assimilation process contains acoustic characteristics that are similar to both the characteristics of the canonical form and the assimilating sound. In case of the coronal place assimilation, this means that the /n/ in *phone box* for example does not completely change to /m/, as some of the acoustic cues to the coronal place of articulation are going to be available even after the assimilation takes place (e.g., Dilley & Pitt, 2007; Gow & McMurray, 2007).

Gow (2002, 2003) argues that the place of articulation of assimilated consonants is recovered through perceptual grouping mechanisms that draw on all available acoustic cues. Through such

mechanisms, listeners associate those acoustic elements that match with the place of articulation features of the upcoming sound to the assimilating segment and those cues that remain are used to recover the place of articulation feature of the assimilated segment. For example, in “cat<sup>p</sup> box”, the surface [labial] feature will be associated with /b/ at the beginning of *box* and the remaining [coronal] feature will be associated with the ambiguous sound at the end of the assimilated word, resulting in identifying the modified word as *cat* rather than *cap*. Crucially, if the same ambiguous sound is followed by a coronal consonant (unviable context for assimilation) rather than a labial consonant, it should be perceived as a labial. In the above example, this would mean that the modified word will be identified as *cap* rather than *cat*. Gow and McMurray (2007) tested this proposal using a version of the visual world paradigm. They presented participants with a set of four pictures on a computer screen. These pictures depicted phrases that contained either a viable or unviable phonological context for triggering assimilation (e.g., “select the *cat*<sup>p</sup> box” or “select the *cat*<sup>p</sup> drawing”, respectively). Importantly, the target words were potentially lexically ambiguous as a result of assimilation (e.g., *cat* vs. *cap*). The authors monitored eye fixations as participants heard the auditory stimuli. Their results indicated more fixations on *cat* compared to *cap* when the following context was viable for place assimilation (i.e., in *cat*<sup>p</sup> box) and to *cap* when the following context was not viable (i.e., in *cat*<sup>p</sup> drawing). These results suggested a compensatory effect as was predicted based on the feature parsing account.

In another experiment, Gow and McMurray (2007) found an anticipatory effect of processing, which they also took as evidence in favor of the feature parsing account. In this case, they presented participants with stimuli that were either carrying assimilatory information (e.g., [n<sup>ɹ</sup>] in *maroon*<sup>ɹ</sup> *goose*), were not carrying any assimilatory information (e.g., the original [n<sup>ɹ</sup>] from *maroon*<sup>ɹ</sup> *goose* was replaced with the [n] from *maroon duck*), or were carrying mismatched acoustic cues (e.g., the [n] from *maroon duck* was replaced with the [n<sup>ɹ</sup>] from *maroon*<sup>ɹ</sup> *goose*). An important factor was that place assimilation could not result in lexical ambiguity of the first word (e.g. *maroon*), however, the following word alternated between a word beginning with [d] (e.g., *duck*) or [g] (e.g., *goose*). The authors found faster fixations on target pictures in the first condition where the stimulus phrase contained anticipatory information consistent with the place of articulation of the following consonant. In other words, after perceiving the final sound of the first word, participants already had an expectation of what was going to be heard next: upon

hearing an assimilated nasal, they were expecting a triggering labial consonant to follow immediately afterward. Conversely, when the sound was not altered as a result of assimilation, listeners did not expect the following word to begin with a sound that would be a viable trigger for assimilation.

The regressive compensation and progressive anticipation effects shown in these experiments are strong indicators of the role of the acoustic cues and the immediate phonological context in processing of assimilation. Critically, the feature parsing account predicts that if assimilation would ever be complete, phonological context could not be used by listeners anymore as there remains no acoustic cue to underlying place of articulation of the assimilated sound and therefore listeners associate the surface acoustic cues of the sound with its underlying place of articulation. This prediction, however, is challenged by recent findings in studies such as Gaskell and Snoeren (2008), discussed in more detail in Section 1.2.1, where perception of completely assimilated lexically ambiguous forms (*run picks* produced as *rum picks*) was found to be modulated by viability of the following phonological context (cf. Gaskell & Marslen-Wilson, 2001).

The feature parsing account also argues that, since general auditory mechanisms are involved in compensation for assimilation, compensation effects should be language-independent and universal (e.g., Gow, 2003; Gow & Im, 2004). For example, Gow and Im (2004) studied Hungarian voicing assimilation and Korean labial-to-velar place assimilation. Korean labial-to-velar assimilation is often argued to be acoustically complete<sup>2</sup> whereas Hungarian voicing assimilation is typically only partial. Gow and Im presented assimilated Hungarian and Korean words to Hungarian and Korean native speakers as well as English speakers who did not have any experience with either of the two types of assimilation. In a series of phoneme detection tasks, they presented VC#CV sequences in different place and voicing assimilation conditions where the following context was either viable or unviable for assimilation to occur. Upon presenting the stimuli to all three groups of listeners, results showed no difference in rate of correct detection based on listeners' language background. This evidence is inconsistent with an account that relies on listeners' phonological knowledge and the results of studies that have

---

<sup>2</sup> Refer to Jun (1996) for an alternative argument.



found language specific effects in resolving phonological variation (e.g., Darcy, Ramus, Christophe, Kinzler & Dupoux, 2009; Mitterer, Kim & Cho, 2013).

### 1.2.3 Perceptual Integration

A third account for how listeners compensate for the effects of place assimilation is the perceptual integration account (Mitterer, Csepe & Blomert, 2006). The perceptual integration account emphasizes language-independent properties of compensation but also assumes a role for language-dependent factors. Specifically, this account proposes that compensation for assimilation occurs at early stages of auditory processing, before the level at which representational features are accessed. This is unlike both inference and feature parsing accounts, both of which assume compensation is taking place at a more abstracted representational level involving phonological features. The proposed account is based on a general concept of perceptual saliency in audition and the interplay between production and perception. When two sounds that are placed in a sequence are acoustically very similar, it is likely that the two sounds are going to be perceived different from one another as a result of perceptual contrast effects (Repp, 1983). Due to the weaker perceptibility of the assimilated sound, and its acoustic similarity to the following segment, acoustic information from the context affects auditory perception of the assimilated sound in accordance with perceptual contrast effects. This would mean that, for example, auditory processing of the acoustic characteristics of an assimilated /n/ (e.g., lowered F2) at the end of *green* in “green<sup>m</sup> box” is influenced by the acoustic characteristics of the following [b]. The outcome of this effect is that the labial cues in [n<sup>m</sup>] will be perceived as being as distinct as possible from labial cues of [b] and therefore closer to the cues to coronal place of articulation to maintain the auditory contrast.

In addition, to account for language-dependent effects of compensation, perceptual integration assumes a role for language-specific phonological knowledge in cases where the underlying form of the assimilated sound remains partially ambiguous, therefore allowing listener’s phonological knowledge to help resolve the ambiguity (Mitterer, Kim & Cho, 2013). Crucially, the perceptual integration account, unlike the feature parsing account, claims no need to rely on the assimilation to be partial and for residual acoustic cues to be present. Instead, it rather assumes that auditory mechanisms act to maximize the contrast between neighboring sounds even when they are almost acoustically identical, as is the case in complete assimilation (Mitterer et al., 2013). It also

differs from the phonological inference account in that it does not assume dependency on language-specific experience as a crucial factor in compensating for assimilation. This is because the proposed level of processing (i.e., the auditory level) is rather early and automatic and therefore the language specific phonological knowledge could only affect perception at later levels of processing, for example, to strengthen an earlier decision made at the auditory level. In their study, Mitterer, Csepe & Blomert (2006) examined the phonological process of liquid assimilation in Hungarian, where /r/ sequences are produced as [rr] across word boundaries, and in Dutch, where liquids do not assimilate in manner in such contexts. In an identification and a discrimination task, they presented listeners from both language groups with words and non-words that ended in either [l] or [r] and followed by a viable or an unviable assimilation context or in isolation. The results revealed that listeners of both language groups were better at identifying the assimilated forms and worse at discriminating an assimilated and an unassimilated canonical form when they were presented with the words in a viable phonological context for assimilation (i.e., the liquid was followed by an /r/). However, some language-dependent effects were also observed in that Hungarian listeners were still generally better than Dutch speakers in categorizing the assimilated sound as /l/, compatible with their phonological knowledge of their native language.

The perceptual integration account stated in Mitterer's earlier work is, however, challenged by the results of a more recent study by Mitterer, Kim & Cho (2013). In this study, the authors found a complete reliance on language-dependent phonological knowledge in compensating for labial-to-velar place assimilation in Korean. As mentioned earlier, Korean labial-to-velar assimilation is an optional but complete assimilation process. Mitterer and colleagues presented native Korean listeners with completely assimilated (ending in a velar stop) or unassimilated (ending in a labial stop) words. They then presented the stimuli in either viable (followed by a k-initial word) or unviable (followed by s-initial word) phonological contexts in an eye tracking experiment. The results showed that Korean listeners looked more often to the target and showed a faster reaction time upon hearing completely assimilated form of the word (ending in a velar consonant) when the phonological context licensed assimilation compared to when it did not. In their second experiment, they again found similar patterns by using a two-alternative forced-choice phonetic categorization task. In this case, the stimuli were similar but the materials were prepared such that the unassimilated and completely assimilated forms were used as the

endpoints on a continuum with six steps in-between, thereby generating different degrees of assimilation. Surprisingly, they found a compensation effect only for completely assimilated forms, indicating that Korean listeners' phonological knowledge regarding the process of place assimilation in this language (i.e., that it is most often complete rather than partial) most prominently affected their processing of the assimilated forms, which better matches with the predictions of the inference account. The perceptual integration account would have predicted a higher rate of compensation for assimilation in acoustically ambiguous cases, where the auditory mechanisms could associate the mixed acoustic information with assimilation at an earlier level. Finally, in a similar categorization task the authors also tested Dutch and English listeners, who were not familiar with labial-to-velar place assimilation in their native language. The compensation effect that was observed in Korean data was no longer found for the other two groups, suggesting that the Korean listeners were in fact using their language-specific phonological knowledge to compensate for assimilation.

### 1.3 Integration of the Contextual Information

Although the major accounts of compensation for assimilation discussed here take distinct perspectives on issues, at certain points their predictions and claims show notable overlap. For example, all three accounts generally agree in that, under certain conditions, phonological context can affect recognition of assimilated words. In addition, all accounts agree to some extent that acoustic cues and phonetic details, specifically the residual acoustic cues associated with the underlying place of articulation of an assimilated sound, can influence compensation for assimilation.

However, the major point of controversy seems to be around the underlying mechanism(s) involved in the integration of contextual information in the process of compensation for assimilation. According to the feature parsing and perceptual integration frameworks, general auditory mechanisms are at play. On the phonological inference account, listeners' language-specific phonological knowledge plays the main role in compensating for assimilation. This controversy remains unresolved, as there is evidence in accordance with both groups of arguments. As noted above, experiments focusing on cross linguistic comparisons have represented one strategy for resolving this debate (Clayards, Niebuhr & Gaskell, 2015; Gow & Im, 2004; Lahiri & Marslen-Wilson, 1991; Mitterer, Kim & Cho, 2013; Skoruppa, Mani &

Peperkamp, 2013). The logic of these studies is built on the idea that general auditory processing and compensation for co-articulation involve language-independent mechanisms whereas probabilistic learning and processes related to phonological knowledge would depend on specific language experience; therefore, the cross-linguistic examination of compensation for assimilation should provide the right tool for understanding the underlying mechanisms. However, one of the main concerns with experiments involving cross-linguistic comparisons is the methodology and in particular, the preparation of stimuli. To interpret the results, the stimuli need to be comparable across the languages being tested. However, phonologically-similar sounds across languages can in fact be acoustically quite distinct. For example, the distribution of stop VOT in Canadian French differs from that of either English or even Continental French (Caramazza and Yeni-Komshian, 1974). Similarly, the main acoustic cue to the vowel [i] is found to be high levels of F3 in French and Swedish, unlike in English, which has been found to be higher levels of F2 (Vaissiere, 2007). These facts make comparing the results of perceptual tests difficult because it is often unclear whether any observed difference is due to the listener's experience with language-specific patterns or simply due to the fact that the listener has encountered a speech sound that is foreign to their ears (cf., Darcy et al., 2009). An alternative method that has been proposed is using listeners' native language sounds. However, that method would not provide a solution to the issue either, because to produce a non-native phonologically altered form, speakers need to mispronounce words, which again makes the stimuli sound unnatural.

One potential way to distinguish automatic auditory processes and probabilistic learning mechanisms is to examine the similarities and differences among what traditionally fall under the same category of phonological processes. Under general auditory accounts, the mechanisms responsible for compensating for place assimilation are expected to be automatic and unconscious and therefore should not show much variation across languages and type of sounds (speech or non-speech). This is not to say that, according to these accounts, the system is not sensitive to variation in acoustic and auditory details across different sounds. Rather, they assume similarity in the general procedures that are involved in processing various inputs.<sup>3</sup> The

---

<sup>3</sup> For example, according to feature parsing account, English listeners can perceive the underlying place of articulation of a labial consonant that has been assimilated to the following coronal consonant through the same mechanisms that they use to compensate for assimilation of a coronal consonant to the following labial consonant,

probabilistic learning account, on the other hand, predicts probabilistic characteristics of sound patterns within individual languages to have an influence on the perceptual processing involved in compensation for place assimilation in those speech sounds. Since, as discussed in earlier sections, place assimilation in nasal and oral stops differ in certain probabilistic aspects, the latter framework, but not the former, would predict differences in the mechanisms involved in perceptual processing of these two sounds when they are assimilated. However, most studies on perceptual processing of place assimilation, particularly those by Gow and colleagues, assume that coronal stop place assimilation and nasal place assimilation, being very similar phonological processes, are dealt with similarly by spoken word recognition mechanisms and therefore they have not tested these two groups of sounds separately. However, as discussed earlier, this assumption should likely be approached with some caution. In two related studies conducted by Mitterer and colleagues (e.g., Mitterer, 2011; Mitterer & McQueen, 2009), it has been claimed that the perception of nasals and stops that undergo the process of place assimilation might involve different processing mechanisms. Mitterer (2011) studied these processes in Dutch in which nasal place assimilation can be triggered when a word ending in /n/ is followed by a word starting with /b/, similar to English (e.g., /tœyn bɑŋk/ → [tœyn<sup>m</sup> bɑŋk] 'garden bench'). Similarly, in words ending in /t/, the process of t-reduction can be triggered when the following word starts with /b/ (e.g. /mɛst bɛstɛlt/ → [mɛsØ bɛstɛlt] 'fertilizer ordered'). Mitterer presented assimilated and reduced forms in viable and unviable contexts in a four-interval oddity-detection task in which four stimuli would be presented to the participants who in turn had to decide whether the second or the third in the set was different from the other three words. For example, for nasal assimilation participants had to discriminate [tœyn.ba] from [tœyn<sup>m</sup>.ba] and [tœyn.stu] from [tœyn<sup>m</sup>.stu]. For stop place assimilation, the pairs to be discriminated were [kyst.bo] versus [kysØ.bo] and [kyst.na] versus [kysØ.na]. Mitterer observed a phonological context viability effect in nasal assimilation but not in t-reduction. However, this difference was observed in an offline discrimination task, which is believed to tap into lower levels of processing. When an online eye tracking task was used instead, the phonological context effect was also observed in

---

even though, the former assimilation (labial-to-coronal) is not attested/common in English (Gow & McMurray, 2007).

stop place assimilation (Mitterer & McQueen, 2009). Therefore, based on these findings, the conclusion was that, unlike in nasal place assimilation, phonological context effects in t-reduction do not arise during early perceptual processing. This could suggest that the phonological processes of nasal place assimilation and stop reduction, even though very similar on the surface, are handled through distinct processing mechanisms. Besides the two studies conducted by Mitterer and colleagues, however, in the majority of studies on compensation for assimilation nasals and stop consonants were often tested together, preventing any variation in their processing to be noticed.

Studying the factors influencing lexical ambiguity resolution can also provide insight into the nature of the processing mechanisms involved in lexical access. Lexical ambiguity that arises as a result of place assimilation makes the recognition of assimilated words even more challenging. For example, in *foam box* [fowmbaks] versus *phone box* [fown<sup>m</sup>baks] or *map box* [mæpbaks] versus *mat box* [mæt<sup>p</sup>baks], after undergoing place assimilation, *foam-phone* or *map-mat* can sound very similar as a result of a strong degree of place assimilation. In such cases, bottom-up information might not be sufficiently helpful with accessing the correct lexical item (e.g., Gaskell, 2003). Under such conditions, any biasing effect of the context must be due to the influence of the higher level information processing. The focus in the current study is therefore on processing mechanisms involved in situations where assimilation can potentially result in lexical competition in similar phonological processes namely nasal and stop place assimilation in English.

## 1.4 Lexical statistics

The effect of lexical factors, such as usage frequency and neighborhood density (phonological, semantic, etc.), on the production and recognition of lexical items has been shown in a number of studies (e.g., Jurafsky, Bell, Gregory & Raymond, 2001; Scarborough, 2004; Segui, Mehler, Frauenfelder & Morton, 1982; Ussishkin & Wedel, 2009; also see Aylett & Turk, 2006, for a review on language redundancy). For example, in words with a higher phonological neighborhood density, namely those words which are phonetically similar to a high number of other lexical items in the lexicon, phonemes were found to be produced more distinctively (e.g., Munson & Solomon, 2004; Wright, 2004). In an analysis of vowel-space in words with different lexical densities, Munson & Solomon (2004) found vowel-space expansion (i.e., further distance

from the average F1 and F2 values) in words with denser neighborhoods. Furthermore, high neighborhood density has been found to result in a lower efficiency in lexical activation (e.g., slower reaction time; Luce, 1986).

Lexical frequency can also influence the production and processing of speech sounds. The probability of reducing speech is often higher in words with higher usage frequency (Bybee, 2001). In addition, more frequent lexical items are often found to be activated and accessed faster and more accurately (e.g., Segui, Mehler, Frauenfelder & Morton, 1982). The frequency of production of a lexical item with a particular phonological variant has also been found to influence ease of lexical access. For example, Ranbom and Connine (2007) conducted a series of priming experiments on the recognition of English words that were produced with a word-medial nasal flap (e.g., *gentle* pronounced as *gennle*). They found faster and more accurate lexical decisions for words that are often produced in English with the flap variant than for words that are less often produced with a flap (i.e., more commonly produced with the canonical /nt/ form). These outcomes suggest that statistics of the lexicon, among other contributing factors, can also influence the patterns of perception and production of lexical items. Ussishkin and Wedel (2002, 2009) proposed that this relationship helps maintain overall contrast among lexical items for ease of access in the absence of a clear phonological contrast (e.g., due to phonological variation). This raises the question regarding the effect of lexical statistics in compensation for place assimilation, especially in cases where the absence of a clear perceptual contrast between the word-final consonants might result in lexical ambiguity (e.g., 'cat box' vs. 'cap box'). In this dissertation, the structure of the English lexicon is examined with respect to the probability of the place assimilation process resulting in lexical ambiguity when compared between words ending in nasal versus oral stop consonants. In particular, it is shown how patterns of the lexicon might reflect patterns of production and perception of assimilated nasals versus stops.

## 1.5 Thesis structure

In previous sections, important aspects of variation and spoken word recognition that were relevant to the current research were outlined. In the following chapters, I present a set of experiments with the aim of investigating these issues further. In Chapter 2, I investigate the acoustic characteristics of assimilated and canonically coronal or labial consonants separately for nasals and oral stops. To this end, I report a series of acoustic analyses on the production data

extracted from the recorded auditory stimuli that are subsequently used in the four main experiments in this thesis. Chapter 3 reports two forced-choice identification experiments in which listeners were required to identify isolated assimilated and unassimilated words ending in nasal or stop consonants. These experiments examined the role of acoustic properties of sounds and residual cues to the underlying place of articulation, independent from the role of the phonological context. The outcomes of these studies also provided a baseline for interpreting the results of the remaining experiments. In Chapter 4, I address the role of phonological context through two experiments using a variant of a priming paradigm, combined with a visual world methodology. The experiments also serve to highlight similarities and differences between the processing of assimilated nasal versus oral stops. In Chapter 5, I discuss the results of a corpus analysis focusing on the structure of the lexicon in which I assess the incidence of words ending in nasal and oral stops whose assimilated forms can entail lexical ambiguity. Finally, in Chapter 6, I provide a summary of the results from the preceding chapters and discuss their implications in relation to the accounts of how listeners manage phonological variation of place assimilation.



## Chapter 2

### Preparation of Auditory Materials and Acoustic Analysis

To investigate the factors that affect recognition of assimilated forms and to compare the processing of assimilation across different natural sound classes, it is crucial to know the degree of assimilation in the sounds being examined. The amount of overlap between the articulators during the production of a sequence of a coronal nasal or oral stop followed by a labial or velar consonant determines the degree of place assimilation. As noted earlier, although place assimilation was traditionally assumed to be complete (i.e., place contrast information would be neutralized), acoustic analyses of assimilated forms have shown that in fact residual cues to the underlying place of articulation of the consonants still exist after undergoing assimilation (e.g., Dilley & Pitt, 2007; Gow, 2002; Gow & McMurray, 2007). An acoustic measure that is usually reported in such studies is formant frequency during the transition from the preceding vowel to the assimilated consonant. For this purpose, usually the second formant (F2) and/or the first or third formants (depending on the place of articulation) are measured at a point around or after the middle of the vowel duration and at the onset of the nasal/stop consonant. For example, Gow (2002) conducted an acoustic analysis on the stimuli he used in a priming experiment consisting of words ending in canonically labial or coronal stop consonants as well as words ending in assimilated coronal stops. Gow measured F2 and F3 values at the points in time corresponding to: 1- the vocalic pitch period with the greatest amplitude usually around the midpoint in the vowel; 2- the penultimate pitch period prior to the stop closure. He subtracted the second measure from the first measure to calculate the amount of formant transition for F2 and F3 in each word. When the average formant change measures were compared across the three sound groups, Gow reported that the average formant frequencies associated with the assimilated coronal stops were always intermediate between the measures corresponding to canonically labial or coronal stops. These differences were significant for F2 transition for both assimilated versus canonically labial (e.g., *right<sup>p</sup>* vs. *ripe*) and assimilated versus canonically coronal comparisons (e.g., *right<sup>p</sup>* vs. *right*); however, F3 differences were only marginally significant when compared between assimilated versus coronal stops and not significant between assimilated versus labial stops. In a later study, Gow and McMurray (2007) conducted acoustic analyses of formant cues on a set of minimal pairs ending in either labial or coronal consonants and they measured all three formant values, including the first formant (F1), prior to stop closure

in their analyses. Words ending in nasals and stops were combined in the analysis. They reported no significant difference among the three types of consonants (labial, coronal and assimilated coronal consonants) for F1 value, and partially significant differences for F3 (between assimilated and labial segments). F2, however, was again significantly different among the three types of segments, suggesting that F2 frequency is an especially informative cue to changes in the place of articulation of coronal consonants.

Dilley and Pitt (2007) also performed an acoustic analysis on words ending in assimilated (or deleted) versus canonically coronal or canonically labial segments from the Buckeye Corpus of Conversational Speech. They controlled for the preceding phonological context by including only the words in which one of the three vowels /ɪ, æ, ε/ preceded the final consonants. They also controlled for the following context by analyzing only the words that were followed by either a labial or a velar word-initial consonant. However, they also grouped nasals and stops together. These authors calculated three measurements involving F2 transition from the preceding vowel, amplitude difference, and consonant duration. The results of their F2 transition analysis revealed no significant difference between canonical labial and assimilated or even deleted forms. However, the F2 difference was significant for assimilated and canonically coronal forms. These results suggest that assimilation was strong/complete in the majority of the cases that were analyzed in the corpus data. The second measure, meaning the amplitude differences for the first two formants, was calculated at two points: (i) at the middle of the preceding vowel; and (ii) close to the preceding vowel endpoint. The results indicated no significant difference among different conditions based on this particular measure. The authors also measured the duration of the word-final segment plus the closure duration of the following word-initial stop. The results of this analysis suggested a significant difference between the three groups of final consonants, with deleted consonants having the shortest duration and the canonically labial or coronal showing the longest duration and the assimilated consonants falling somewhere in between the two groups. Based on these findings, F2 transition was again shown to be a strong indicator of variations in the place of articulation of nasal and stop consonants. In addition, changes in consonant duration clearly correspond with different types of modifications, such as assimilation or deletion, and therefore could be understood as another important perceptual cue for differentiating between the canonical and modified forms.

In the four experiments reported in the current thesis, the process of compensation for place assimilation in nasal and oral stops is the focus of the investigation. Of interest is whether and how listeners compensate for place assimilation when they are presented with naturally produced assimilated words in presence or absence of the triggering phonological context for assimilation. For this purpose, a priming paradigm combined with an eye tracking methodology (only in Experiments 3 & 4) and a forced choice identification task was used on naturally produced stimuli that end in unassimilated or assimilated nasal stops (Experiments 1 & 3) or oral stops (Experiments 2 & 4). In each experimental trial, participants heard and identified an unassimilated or assimilated prime word followed by either an unassimilated or assimilated target word. In Experiments 1 and 2, the prime and the target words were presented in isolation (i.e., the triggering context for assimilation was not provided) and in Experiments 3 and 4 all stimuli items were presented in the original carrier sentences that provided the phonological context for assimilation.

In the next section, I describe the general methodology for production and preparation of the auditory materials used in the four experiments that follow. I then present and discuss the results of a series of acoustic analyses on these stimuli involving words ending in assimilated coronals, and canonically coronal and canonically labial consonants. The results of the analyses are reported separately for nasal and stop consonants and are used for examination and validation of the stimuli in terms of the degree of assimilation. This section also provides the necessary information for interpreting the results of the main experiments that follow.

## 2.1 Preparation of Auditory Materials

Two lists, each consisting of 48 monosyllabic English words, were prepared to be used as the critical stimuli in Experiments 1-4. In one list, all words ended in a coronal nasal /n/ (Expts. 1 & 3, Appendix A) and in the second list, all words ended in a coronal stop consonant /t/ (Expts. 2 & 4, Appendix B). Since a priming paradigm was used in all four experiments, half the words in each list were selected as prime words and the other half as targets. To foreshadow the experiments, the prime items were included to test if the recent act of compensating for assimilation can facilitate a similar process during the comprehension of the target items (see Chapter 3 for an overview of the priming paradigm). All target words were potentially lexically ambiguous when they underwent place assimilation (e.g., *line* can perceptually become similar to

*lime*). From the list of prime words, eight were potentially lexically ambiguous when assimilated to a labial place of articulation (e.g., *scan-scam*), whereas the remaining 16 words were unambiguous (e.g., *lean* where there is no word *leam*).<sup>4</sup> All words (except for the words *ate* and *screen*) had the syllable structure of CVC or CCVC that would allow for a vowel to always precede the final consonant. The words were selected from various lexical categories such as verbs, nouns, prepositions or adjectives. This made it possible to test a larger list of words in the current experiments compared to previous studies (e.g., only 16 words were used in Gow & McMurray, 2007).

Prime words were recorded within a carrier sentence in form of the instruction “Click on the \_\_\_ button” and the carrier sentence “Now click on the \_\_\_ button” was used for target words. A female Canadian English native speaker (Toronto accent) with knowledge of English phonetics and phonology was recorded producing all the auditory experimental materials. The speaker was recorded producing two versions of each sentence. In the first version, a careful pronunciation of the sentences resulted in the production of words in which the final consonant was not altered by assimilation. In the second version a casual pronunciation of the sentences was adopted to produce assimilated forms. All stimuli were recorded in a sound attenuated booth at University of Toronto Phonetics Lab. The sentences were recorded using a DPA 4011 cardioid microphone on a Sound Device 722 digital recorder with a sampling frequency of 44.1 KHz and 24-bit resolution. Throughout the session, the speaker’s distance from the microphone was at about 30 cm. Several recordings of each instruction were made for each version (assimilated/unassimilated). Subsequently, the most natural-sounding tokens were selected. The decisions were based on the author’s own perceptual judgment which was then followed by waveform and spectrogram analysis of the sound files. The intensity level of selected tokens was adjusted to maintain an average amplitude level of 70 dB across the duration of utterance. All acoustic measurements and manipulations were conducted using Praat software (Version 5.3.23).

In addition, 72 word pairs were also recorded to be used in filler trials to counteract strategic expectations and disguise the manipulation of interest (Appendix C). The auditory stimuli for the

---

<sup>4</sup> The asymmetry in the number of lexically ambiguous versus unambiguous words in the set of prime words is due to the limited number of available familiar (non-archaic) words that could be selected from the English lexical inventory.

filler trials consisted of words that did not end in coronal nasal or oral stop consonants and had either semantic or phonological relationships that were not the same as in the critical trials. In twelve filler prime-target word pairs, either or both words ended in a labial nasal (Exp. 1 & 3; e.g., *shame*) or a labial stop (Expts. 2 & 4; e.g., *ship*). This set of filler items served the purpose of neutralizing participants' potential expectation that all critical items would end in a coronal consonant.

## 2.2 Acoustic Analysis

To examine the degree of assimilation and establish the overall validity of stimuli in terms of assimilatory features, a set of acoustic analyses were conducted on the recordings. Because the assimilated/unassimilated stimuli were produced naturally, it was crucial to analyse the acoustic cues to the place of articulation of the word-final consonants to validate the presence of assimilation (i.e., labial cues) in the assimilated words. Also, because the prime words were never used as the target words and vice versa, one additional concern was to ensure the assimilatory characteristics of the prime and target words were in fact comparable across the two word sets. As mentioned earlier, the purpose of using a priming paradigm in the experiments was to test whether a recent act of “undoing” assimilation could in fact facilitate processing of assimilation in a subsequent target word. If the assimilated primes show significantly different acoustic characteristics compared to assimilated targets, then that might affect the activation of the target words in the following sentence and any effect of priming. Therefore, acoustic measures associated with nasal and stop consonants' place of articulation, including three measures of formant frequency as well as consonant duration (only for stop consonants), were calculated and reported for both prime and target words in the sections that follow.

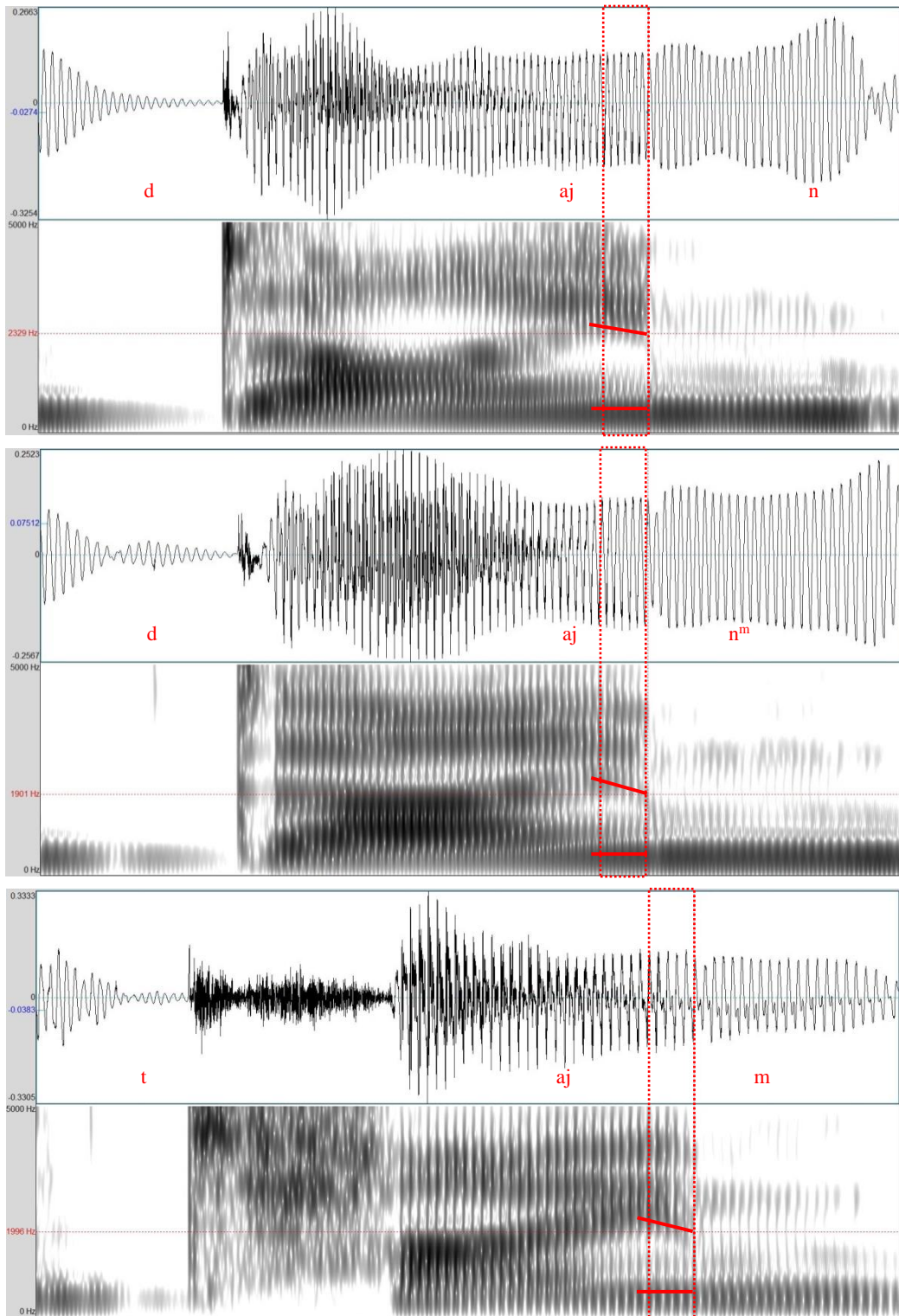
### 2.2.1 Formant Frequency

F2 frequency of vowels is reported to be especially affected by place of articulation features of the surrounding consonants (e.g., Dilley & Pitt, 2007; Gow, 2002; Hon, 2005). The F2 of the preceding vowel around the onset of the word-final nasal or oral stop tends to be lower when the following consonant carries labial features, such as in assimilated or labial consonants, compared to when the following sound is a coronal. Therefore, to test for the validity of the experimental stimuli and the degree of assimilation, three formant frequency measures were analyzed and compared in prime and target words ending in assimilated and unassimilated coronal nasal and

oral stops: (i) F2 frequency at the offset of the vowel/onset of the word-final consonants; (ii) slope of F2 transition from vowel to the final consonant; and (iii) the difference between F2 and F1 frequencies at the offset of the vowel. The first measure, raw F2 value, was also used in the study conducted by Gow and McMurray (2007), where they analyzed F1, F2 and F3 data and found F2 values to be the most relevant in distinguishing assimilated versus canonically labial or coronal stop consonants. The second measure, F2 slope/F2 transition, is the acoustic characteristic that has been commonly reported in previous studies investigating acoustic characteristics of assimilated coronal versus unassimilated coronal or labial stop consonants (Dilley & Pitt, 2007; Gow, 2002). In the current study, another measure, namely F2 and F1 difference, is also analyzed for an additional relative measure of F2 frequency. For the experimental stimuli to be validated, formant frequency measures are expected to be comparable across prime and target words.

#### 2.2.1.1 Nasal consonants

A series of spectrograms for the words *dine* [dajn], *dine* [dajn<sup>m</sup>] and *time* [tajm] are presented in Figure 2-1 respectively. The second and third spectrograms show a considerable drop in the second formant frequencies (F2) around the onset of the word-final assimilated or labial nasal consonants.



**Figure 2-1** Sample waveforms and corresponding spectrograms for *dine* [dajn] (upper panel), *dine* [dajn<sup>m</sup>] (middle panel) and *time* [tajm] (lower panel). Solid red lines show the first and second formants as they transition within the analysis window (red rectangles) from 20 ms prior to the offset of the vowel [aj] to the onset of the following nasal.

The first measurement was the raw value of the second formant (F2) measured at the end of the vowel/onset of the nasal murmur. The average F2 values for assimilated and unassimilated coronal nasals for prime or target words are shown in Table 2-1. The results follow the predicted pattern: The average F2 value for assimilated nasals was lower than the average F2 for the unassimilated coronal nasals in both prime and target sentences. The results of a paired t-test indicated that this difference was in fact significant for both groups of primes and targets (primes:  $t(23) = 3.36$ ,  $p = 0.002$ ; targets:  $t(23) = 2.86$ ,  $p = 0.008$ ). A set of two sample t-tests on the other hand indicated the average F2 value was not significantly different between prime and target words in either assimilated ( $t(46) = 0.21$ ,  $p = 0.83$ ) or unassimilated coronal ( $t(46) = 0.47$ ,  $p = 0.63$ ) conditions. In addition, the amount of difference between assimilated and unassimilated conditions in F2 value (i.e., the magnitude of F2 change) did not show a significant difference when compared between prime and target words ( $t(46) = 0.95$ ,  $p = 0.34$ ), which further confirms the comparability of the chosen stimuli within and across prime and target groups.

**Table 2-1** Three measures of F2 frequency and the corresponding standard deviations (SD) for prime and target words ending in nasal consonants.

Measure	Condition	Prime (N = 24)		Target (N = 24)	
		Mean (Hz)	SD	Mean (Hz)	SD
<b>F2</b>	unassimilated	1929.1	451.2	1995.8	510.3
	assimilated	1773.3	470.7	1745.2	433.9
<b>F2 slope</b>	unassimilated	-93.9	169.7	-74.2	156.9
	assimilated	-202.2	142.2	-228.7	197.2
<b>F2 – F1</b>	unassimilated	1464.3	479.7	1428.7	602.1
	assimilated	1296.2	504.3	1139	564.5



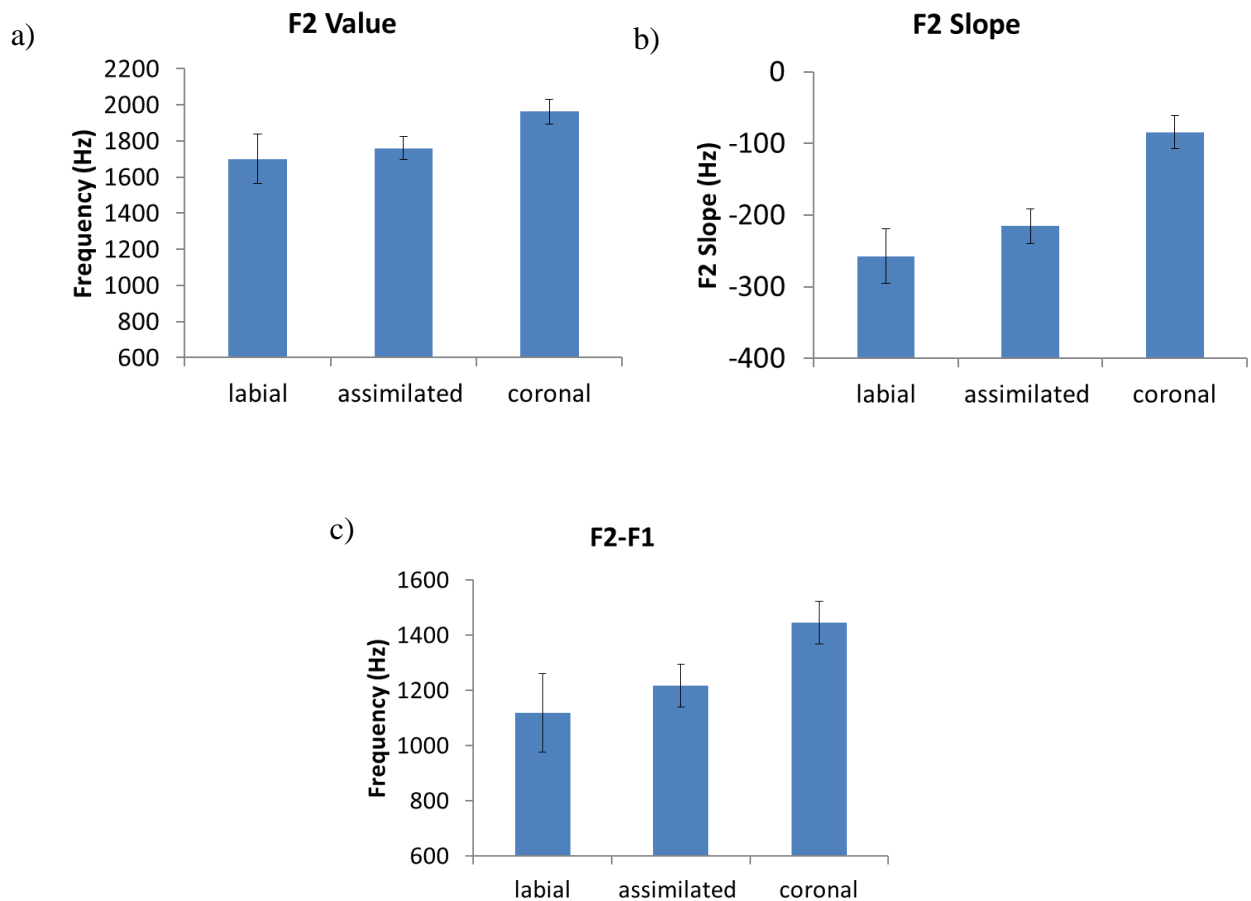
Because the vowel nuclei were not the same across the stimuli, a relative rather than an absolute measure of F2 value would better reflect the patterns of formant transition. Therefore, the second measure was F2 transition slope that was calculated based on the value of F2 at the nasal onset relative to the F2 value at 20 ms prior to the nasal onset during the steady state of the vowel. This slope is expected to show a higher amount of decrease in the nasals with labial features compared to the unassimilated coronal nasals. The result of this calculation was again as expected, with on average a steeper falling F2 slope for assimilated nasals ( $M = 215$  Hz) compared to unassimilated coronal nasals ( $M = 84$  Hz). Paired t-tests indicated significant differences between the assimilated versus unassimilated coronal forms in both groups (primes:  $t(23) = 2.62, p = 0.01$ ; targets:  $t(23) = 4.17, p < 0.001$ ). Two-sample t-test results also showed no significant difference between primes and targets in either assimilated ( $t(46) = 0.53, p = 0.59$ ) or unassimilated coronal ( $t(46) = 0.41, p = 0.67$ ) conditions. The magnitude of the slope change (i.e., the amount of the difference between assimilated and unassimilated primes or targets) was also not significantly different between the prime and target words ( $t(46) = 0.95, p = 0.34$ ).

Finally, the third measurement was the difference between F2 and F1 frequencies at the offset of the vowel/nasal onset time, which is also a relative measure of formant value. Since the F2 of the vowel tends to decrease around the onset of the following assimilated or labial nasal and at the same time F1 value remains relatively stable, a smaller amount of difference between F2 and F1 frequencies at that point in time is expected for assimilated and labial nasals compared to unassimilated coronals. For the current stimuli in fact this was the case. The results of the paired t-tests were found to be significant for both groups of prime and target words (primes:  $t(23) = 3.67, p = 0.001$ ; targets:  $t(23) = 2.96, p = 0.006$ ). Two-sample t-test results suggested no significant difference between prime and target groups in either of the assimilated ( $t(46) = 1.01, p = 0.31$ ) or unassimilated coronal ( $t(46) = 0.22, p = 0.82$ ) conditions and the magnitude of the difference between unassimilated and assimilated words for this measure was again not significantly different between primes and targets ( $t(46) = 1.16, p = 0.25$ ).<sup>5</sup>

---

<sup>5</sup> Three sets of two-way between-item ANOVA, with the variables of Word Type (prime vs. target) and Assimilation (unassimilated vs. assimilated), were also conducted on the data from the three acoustic measurements. The results of the tests conformed to the results of the t-tests. The effect of Assimilation was significant based on all measurements (F2:  $F(1, 92) = 4.53, p = 0.03$ ; F2 slope:  $F(1, 92) = 14.7, p < 0.001$ ; F2-F1:  $F(1, 92) = 4.31, p = 0.04$ ). No significant main effect of Sentence Type (F2:  $p = 0.8$ ; F2 slope:  $p = 0.9$ ; F2-F1:  $p = 0.4$ ) and no interaction effect between Assimilation and Sentence Type (F2:  $p = 0.6$ ; F2 slope:  $p = 0.5$ ; F2-F1:  $p = 0.6$ ) were observed.

As an additional examination of the extent of assimilation present in the experimental stimuli, the same measures were also calculated for a small separate sample of words ending in labial nasals (15 words from the filler items) and compared with the corresponding measures for assimilated and unassimilated coronal words from both prime and target lists. The graphs in Figure 2-2 represent the results of the three measures of average F2 value (Fig. 2-2a), F2 transition slope (Fig. 2-2b) and the difference between F2 and F1 values (Fig. 2-2c).



**Figure 2-2.** F2 frequency measures for word-final labial, coronal and assimilated nasals. The measures were averaged across prime and target words.

As shown in Figure 2-2, assimilated nasals tended to pattern more closely with the labial nasals rather than the coronal nasals based on the three measures of formant frequencies. A linear mixed effects model using *lmerTest* package (V 2.0-32, Kuznetsova, Brockhoff, & Christensen, 2016) in R (V 3.2.5, R Core Team, 2016) was conducted on the data corresponding to the three

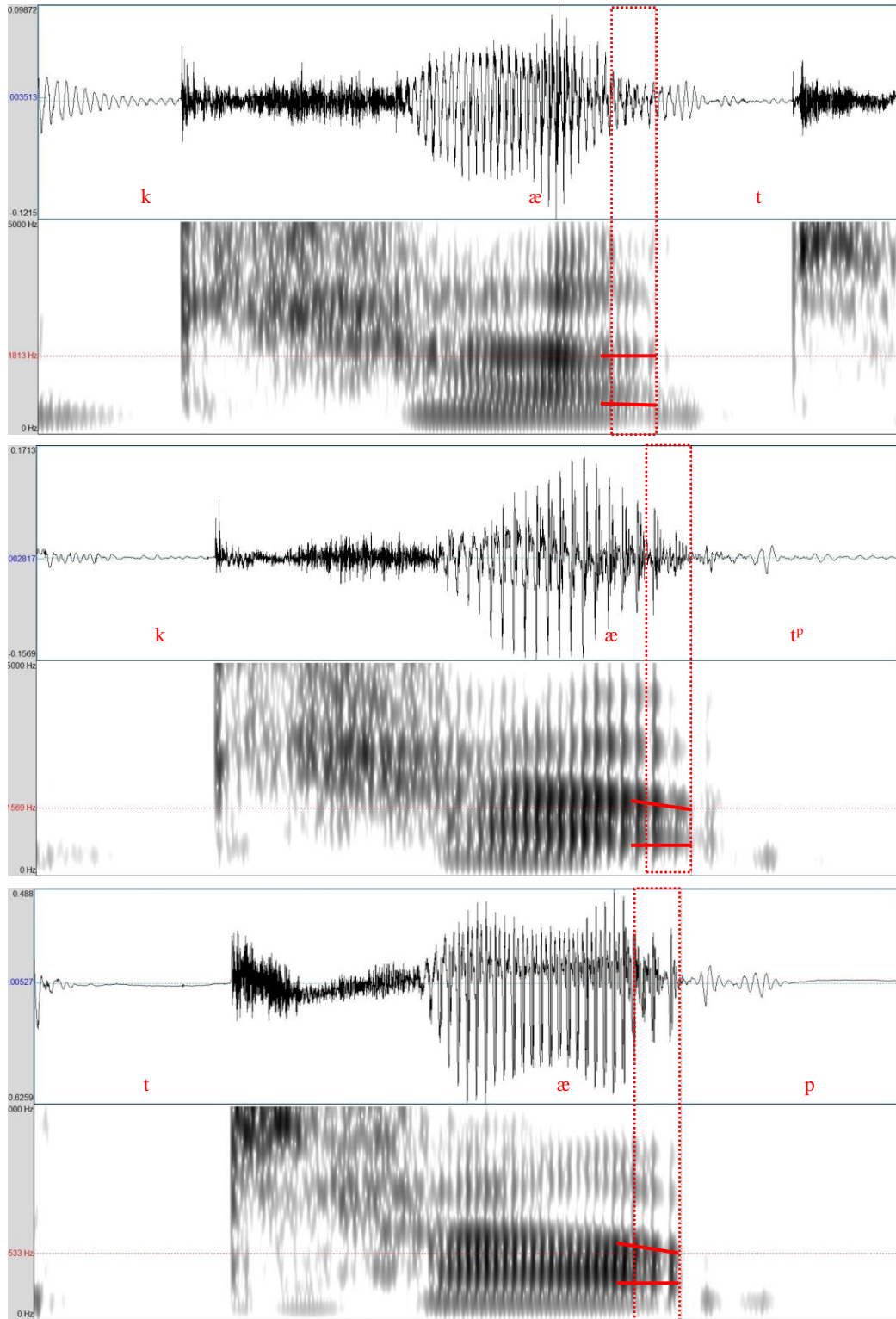
measures. A fixed effect for consonant (unassimilated coronal, labial, assimilated coronal) and a random intercept for item (word token) were included in the model. The results of the analysis on the absolute F2 value data showed a marginal difference between labial and unassimilated coronal nasals ( $M = 1700$  vs.  $1963$  Hz;  $\beta = 262.65$ ,  $SE = 140.8$ ,  $t = 1.87$ ,  $p = 0.06$ ) and no significant difference between labial and assimilated nasals ( $M = 1700$  vs.  $1759$  Hz;  $\beta = 59.42$ ,  $SE = 140.8$ ,  $t = 0.42$ ,  $p = 0.67$ ). The results also revealed significantly smaller difference between F2 and F1 for labial nasals compared to unassimilated coronal nasals ( $M = 1120$  vs.  $1447$  Hz;  $\beta = 326.72$ ,  $SE = 159.76$ ,  $t = 2.05$ ,  $p = 0.04$ ) but no significant difference was found between labial and assimilated coronal nasals ( $M = 1120$  vs.  $1218$  Hz;  $\beta = 97.87$ ,  $SE = 159.76$ ,  $t = 0.61$ ,  $p = 0.54$ ). Finally, the results of the analysis on F2 slope data showed a significant difference between labial and unassimilated coronal nasals ( $M = -257$  vs.  $-84$  Hz;  $\beta = 172.94$ ,  $SE = 48.46$ ,  $t = 3.57$ ,  $p < 0.001$ ) but the difference between labial and assimilated nasals was not found to be significant ( $M = -257$  vs.  $-215$  Hz;  $\beta = 41.61$ ,  $SE = 48.46$ ,  $t = 0.86$ ,  $p = 0.39$ ). These results suggest a strong acoustic similarity between labial and assimilated word-final nasal consonants in the words used as stimuli in the current study. Despite the outcome of the statistical tests, the means for the labial and assimilated coronals are not strictly the same (and the latter tend to fall “in-between” unassimilated coronals and labials). This may mean there are some residual cues to place of articulation in the assimilated coronal nasals.<sup>6</sup>

### 2.2.1.2 Stop Consonants

The spectrograms in Figure 2-3 correspond to the words *cat* [kæt], *cat* [kæt<sup>p</sup>] and *tap* [tæp] respectively. The decrease in the value of the second formant around the onset of the word-final assimilated or labial oral stop consonants can be observed in the second and third spectrogram in comparison with the first spectrogram.

---

<sup>6</sup> The results of the statistical comparisons between labial nasals with unassimilated coronal or assimilated coronal nasals should be interpreted with caution because there is a difference in sample size (i.e., smaller number of words ending in labial nasals) that might have an effect on the statistical power.



**Figure 2-3** Sample waveforms and corresponding spectrograms for *cat* [kæt] (upper panel), *cap* [kætp] (middle panel) and *tap* [tæp] (lower panel). Solid red lines show the first and second formants as they transition within the analysis window (red rectangle) from 20 ms prior to the offset of the vowel [æ] to the onset of the following stop consonants.

To validate the intended characteristics of the experimental stimuli, the same three formant transition measures reported for the nasal data set were also examined for the stimuli containing stop consonants and were compared between assimilated and unassimilated prime and target words. Table 2-2 provides a summary of the results. The first measure, the average F2 value at the offset of the vowel/stop onset time, is expected to be smaller in prime and target words ending in assimilated coronal stops compared to words ending in unassimilated coronal stops (e.g., *hot<sup>p</sup>* button vs. *hot* button). A comparison between assimilated and unassimilated words shows that F2 behaves according to this prediction. The results of a paired t-test indicated a significant difference between assimilated and unassimilated forms in both groups of primes and targets in the expected direction (primes:  $t(23) = 7.3, p < 0.001$ ; targets:  $t(23) = 5.2, p < 0.001$ ). In addition, a series of two-sample t-tests using the raw value of F2 as the dependent measure showed no significant difference between prime and target words, in either assimilated ( $t(46) = 0.9, p = 0.36$ ) or unassimilated coronal ( $t(46) = 1.3, p = 0.2$ ) conditions. Further, the magnitude of the difference between assimilated and unassimilated coronal consonants in F2 value was not significant when compared between primes and target words ( $t(46) = 0.18, p = 0.85$ ).

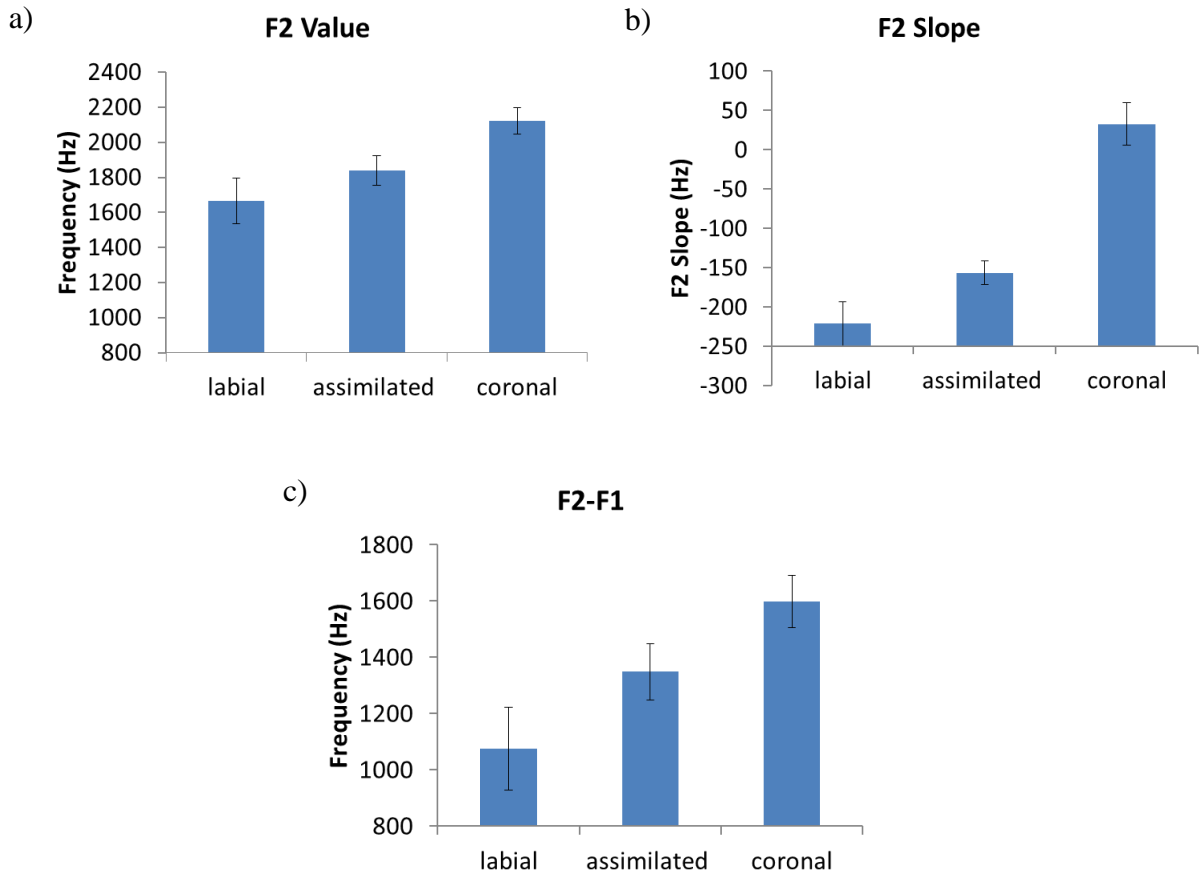
The slope of the F2 transition was examined by calculating the difference between the F2 value at the onset of the stop and 20 ms prior to stop onset, where the vowel transitions into the stop closure. Here, the expected pattern is a steeper slope in assimilated forms compared to unassimilated forms (i.e., coronals). The result of this calculation was as expected, with the F2 slope in assimilated stop consonants showing a 103 Hz more decrease relative to unassimilated coronal stop consonants. Paired t-tests using F2 transition slope as the dependent measure indicated significant differences between the assimilated versus unassimilated coronal forms in both groups (primes:  $t(23) = 5.6, p < 0.001$ ; targets:  $t(23) = 3.2, p = 0.003$ ). Two-sample t-tests did not show any significant differences across the prime and target word lists in either of the assimilation conditions (assimilated:  $t(46) = 0.57, p = 0.57$ ; unassimilated:  $t(46) = 1.1, p = 0.28$ ) and there was no significant difference in the magnitude of F2 slope change between unassimilated coronal stops and assimilated tokens across prime and target words ( $t(46) = 1.45, p = 0.15$ ).

**Table 2-2** Three measures of F2 frequency and the corresponding standard deviations (SD) for prime and target words ending in stop consonants.

Measure	Condition	Prime (N = 24)		Target (N = 24)	
		Mean (Hz)	SD	Mean (Hz)	SD
<b>F2</b>	unassimilated	2026.4	513	2221.4	527.2
	assimilated	1764.7	524.7	1917.5	635.8
<b>F2 slope</b>	unassimilated	62.3	169.5	2.7	204.7
	assimilated	-165.4	85.2	-147.8	124.3
<b>F2 – F1</b>	unassimilated	1483.8	617.7	1711.8	650.1
	assimilated	1237.3	618.7	1458.1	740.5

Finally, the difference between F2 and F1 values at the stop onset also matched the expected pattern: assimilated forms showed a smaller amount of difference compared to unassimilated forms. The results of the paired t-tests were found to be significant in the expected direction for both groups of prime and target words (primes:  $t(23) = 5.8$ ,  $p < 0.001$ ; targets:  $t(23) = 4.5$ ,  $p < 0.001$ ). Two-sample t-tests conducted on the difference between F2 and F1 across the prime and target word lists showed no significant effects in either assimilated ( $t(46) = 1.12$ ,  $p = 0.27$ ) or unassimilated coronal ( $t(46) = 1.26$ ,  $p = 0.22$ ) conditions, and the magnitude of the change between assimilated and unassimilated conditions was again not significantly different between prime and target words ( $t(46) = 0.58$ ,  $p = 0.56$ ).<sup>7</sup>

<sup>7</sup> Two-way between-item ANOVAs, with the variables of Sentence Type (prime vs. target) and Assimilation (unassimilated vs. assimilated), were conducted on the acoustic data. The results of the tests and the results of the t-tests agreed for most measures in that the effect of Assimilation was significant except for the measure F2 and F1 difference (F2:  $F(1, 92) = 6.29$ ,  $p = 0.04$ ; F2 slope:  $F(1, 92) = 36.75$ ,  $p < 0.001$ ; F2-F1:  $F(1, 92) = 3.46$ ,  $p = 0.07$ ).



**Figure 2-4** F2 frequency measures for word-final labial, coronal and assimilated stops. The measures were averaged across prime and target words.

The same formant transition measures were used to examine differences between oral stops in the assimilated and unassimilated coronal conditions (e.g., /t/ in *cat* versus *cat<sup>p</sup>*) and a small sample of the filler words that ended in labial oral stops (e.g, *tap*; 13 words). Figure 2-4 depicts the patterning of the three types of stop consonants with respect to the three measurements. The overall pattern suggests that the assimilated forms are intermediate between the unassimilated

---

No significant main effect of Sentence Type (F2:  $p = 0.1$ ; F2 slope:  $p = 0.5$ ; F2-F1:  $p = 0.1$ ) and no interaction effect between Assimilation and Sentence Type (F2:  $p = 0.9$ ; F2 slope:  $p = 0.2$ ; F2-F1:  $p = 0.9$ ) were observed.

coronal and unassimilated labial consonants for each of the dependent variables. This is similar to the pattern found in Gow and McMurray (2007), where the presence of intermediate cues between labial and coronal place of articulation in assimilated consonants was taken as evidence that sub-phonemic cues are available during speech perception and word recognition. A mixed effects model using *lmer* in R was conducted on the data corresponding to the three measures of raw F2 value at the onset of the nasal consonants, the difference between F2 and F1, and F2 slope. The model included consonant type (coronal, labial, or assimilated) as the fixed effect and a random intercept for item. The results showed that the absolute F2 is significantly different between labial and coronal stops ( $M = 1665$  vs.  $2124$  Hz;  $\beta = 459.24$ ,  $SE = 168.34$ ,  $t = 2.73$ ,  $p = 0.008$ ). However, on average F2 was not significantly different between labial and assimilated stops ( $M = 1665$  vs.  $1841$  Hz;  $\beta = 176.47$ ,  $SE = 168.34$ ,  $t = 1.05$ ,  $p = 0.3$ ). The results also showed a significant difference between labial and coronal stops when compared based on the amount of F2 and F1 difference ( $M = 1074$  vs.  $1598$  Hz;  $\beta = 523.67$ ,  $SE = 199.65$ ,  $t = 2.62$ ,  $p = 0.01$ ), but again the difference between labial and assimilated stops was not found to be significant ( $M = 1074$  vs.  $1348$  Hz;  $\beta = 273.65$ ,  $SE = 199.65$ ,  $t = 1.37$ ,  $p = 0.18$ ). Finally, the results revealed a significant difference between labial and coronal stops in the degree of F2 slope ( $M = -221$  vs.  $-32$  Hz;  $\beta = 253.86$ ,  $SE = 46.19$ ,  $t = 5.5$ ,  $p < 0.001$ ) but the difference between labial and assimilated nasals was not found to be significant ( $M = -221$  vs.  $-157$  Hz;  $\beta = 64.82$ ,  $SE = 46.19$ ,  $t = 1.4$ ,  $p = 0.16$ ). These results suggest that, based on acoustic measures, the degree of assimilation of the word-final stop consonants in the set of stimuli used in the current study is strong. However, the average values corresponding to the labial and assimilated coronal stop consonants, depicted in Figure 2-4, indicates that these measurements are not exactly the same (assimilated coronals are “in between” the other two groups) and therefore there may be residual acoustic cues to the underlying place of articulation of the assimilated coronal stop consonants.<sup>8</sup>

---

<sup>8</sup> The results of the statistical comparisons between labial stops with unassimilated coronal or assimilated coronal stops should be interpreted with caution because of the smaller number of words ending in labial nasals compared to the number of words from the other two groups.



### 2.2.2 Consonant Duration

In addition to formant transition cues, which are often considered the most relevant cues to place of articulation of the consonants, consonant duration is also found to be important in stop consonants (Dilley & Pitt, 2007). Since assimilation in stops often involves lack of closure release (for oral stops) and shortened gap between the assimilated and the following assimilating consonants, the duration of stops is on average shorter when assimilated than when unassimilated.

To examine the experimental stimuli for any variation in duration, the onset of the nasal or oral stop closure (/n/ or /t/) and the offset of the following labial stop /b/ at the beginning of the following word *button* were measured by inspecting the waveform and spectrogram of the sounds. Due to lack of release burst in assimilated oral stop consonants, it was not possible to accurately mark the exact boundary of the word-final oral stops and the following /b/ in this condition; therefore, following Dilley and Pitt (2007), the combined duration of the stop and the following context (i.e., /tb/), rather than duration of isolated stops, was calculated. For comparability, duration of nasal stops was also calculated for the same interval. Closure duration was therefore measured by subtracting the onset of /n/ or /t/ from the offset of /b/ marked at the end of the stop release. Table 2-3 provides the mean duration of the /tb/ and /nb/ in assimilated and unassimilated prime and target words.

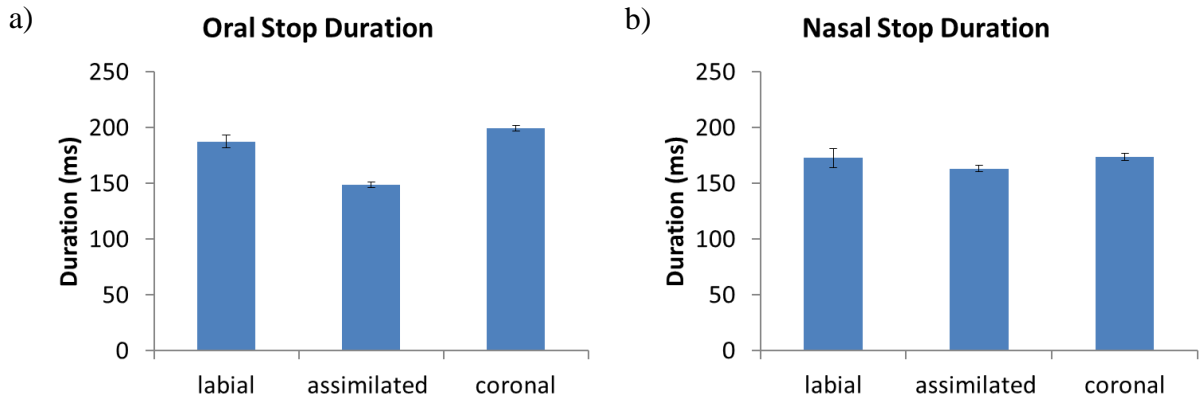
A visual inspection of the duration of the consonants in Table 2-3 indicates that, in the current set of stimuli, assimilated oral and nasal stops are on average shorter than unassimilated stops. A paired t-test between assimilated and unassimilated coronal nasal stops, with consonant duration as the dependent measure, showed that this comparison was significant for target words ( $t(23) = -3.57, p = 0.002$ ), however, the difference failed to reach significance for prime words ( $t(23) = -1.53, p = 0.14$ ). A similar set of comparisons between assimilated and unassimilated coronal oral stops also showed significant differences for both prime ( $t(23) = -11.8, p < 0.001$ ) and target words ( $t(23) = -11.19, p < 0.001$ ). A series of two-sample t-tests on the same measurement showed a significant difference between the unassimilated prime and unassimilated target words for both nasals ( $t(46) = 2.89, p = 0.005$ ) and oral stops ( $t(46) = 3.04, p = 0.004$ ). However, the difference between assimilated coronal prime and assimilated coronal target words was not significant for either nasal stops ( $t(46) = 1.61, p = 0.11$ ) or oral stops ( $t(46) = 1.51, p = 0.14$ ).

Also, the magnitude of the difference in duration between assimilated and unassimilated forms, when compared between prime and target words, did not reach significance for either nasal stops ( $t(46) = 1.43, p = 0.16$ ) or oral stops ( $t(46) = 1.17, p = 0.25$ ). This relative measurement of duration shows that, despite the difference between unassimilated prime and target words in their raw duration values, the amount of change when the consonants become assimilated is not significantly different for prime words compared to target words.<sup>9</sup>

**Table 2-3** Average combined duration of the word-final oral and nasal stops and the following labial stop and the corresponding standard deviation (SD) for prime and target words ending in stop consonants.

Measure	Consonant	Condition	Prime (N = 24)		Target (N = 24)	
			Mean (ms)	SD	Mean (ms)	SD
<b>C-Duration</b>	Oral stop	unassimilated	206.9	20.5	191.8	12.9
		assimilated	152.7	17.3	144.9	18.3
	Nasal stop	unassimilated	164.9	22	182.1	19.1
		assimilated	158.8	19.7	167.9	19.3

<sup>9</sup> Two two-way between-item ANOVAs, with the variables of Word Type (prime vs. target) and Assimilation (unassimilated vs. assimilated), were conducted on consonant duration data for nasal and oral stops. The results overall matched with the results of the t-tests in that, for both nasal and oral stops, there were significant main effects of Sentence Type (nasal stops:  $F(1, 92) = 10.28, p = 0.002$ ; oral stops:  $F(1, 92) = 10.22, p = 0.002$ ) and Assimilation (nasal stops:  $F(1, 92) = 6.17, p = 0.01$ ; oral stops:  $F(1, 92) = 200.12, p < 0.001$ ). There were no significant interaction effects between Assimilation and Sentence Type for either nasal stops ( $p = 0.3$ ) or oral stops ( $p = 0.3$ ).



**Figure 2-5** Average combined duration of the word-final stop consonants and the following labial stop for labial, assimilated coronal and unassimilated coronal oral stops (left panel) and nasal stops (right panel). The data was averaged across prime and target words.

Finally, Figure 2-5 illustrates the duration measurements corresponding to labial stops retrieved from the sample filler items for oral stops (left panel) and nasal stops (right panel). A visual inspection of the average duration of labial stops and unassimilated coronal stops (combined prime and target words) indicates that they are very similar (orals stops: 188 vs. 199 ms; nasal stops: 173 vs. 174 ms). However, assimilated coronal stops were shorter in duration than either unassimilated labial and coronal stops in both groups (orals stops: 149 ms; nasal stops: 163 ms). Two mixed effects analyses (*lmer* test in R), with consonant type (unassimilated coronal, labial, assimilated coronal) as the fixed effect, were conducted on the data. Random intercept for item was included in the models. The results showed no significant difference between duration of labial and unassimilated coronal nasals ( $M = 173$  vs.  $174$  ms;  $\beta = -9.36$ ,  $SE = 7.05$ ,  $t = -1.33$ ,  $p = 0.19$ ) or between duration of labial and assimilated coronal nasals ( $M = 173$  vs.  $163$  ms;  $\beta = 0.81$ ,  $SE = 7.05$ ,  $t = 0.12$ ,  $p = 0.91$ ). The results of the analysis for oral stops however showed a marginally significant difference between the duration of the labial stops compared to the unassimilated coronal stops ( $M = 188$  vs.  $199$  ms;  $\beta = 11.72$ ,  $SE = 5.85$ ,  $t = 2.0$ ,  $p = 0.05$ ) and a highly significant difference when the duration was compared between labial and assimilated coronal stops ( $M = 188$  vs.  $149$  Hz;  $\beta = -38.88$ ,  $SE = 5.85$ ,  $t = -6.65$ ,  $p < 0.001$ ). This pattern of results in fact matches the predictions that lack of release burst and shorter pauses/faster

articulation of the assimilated oral stops would naturally result in an overall shorter duration of the consonant.

## 2.3 Discussion

The purpose of the acoustic analyses reported in this chapter was to examine the degree of assimilation in the recorded stimuli as a means of validating their use in the four experiments that follow. Three cues to F2 transitions (from the preceding vowel to the word-final stop or nasal consonants) were measured and compared across assimilated and unassimilated conditions for both prime and target carrier sentences. In addition, for the final oral stops, consonant duration was calculated and compared across the various conditions.

The results showed that the two groups of prime and target words are comparable with respect to all measured acoustic cues in both assimilated and unassimilated conditions, for both nasals as well as oral stop consonants. The exception to this general pattern was consonant duration, which showed a variation presumably depending on the duration of the prime and target carrier sentences. Therefore, based on the current patterns of results, it would be unlikely that any potential effect of phonological context and/or priming in the subsequently reported experiments would stem from uncontrolled differences in the assimilatory characteristics of the prime or target words.

Moreover, the overall results indicated that all measured acoustic cues for place of articulation in the assimilated condition were significantly different from their corresponding measures in the unassimilated coronal condition, and that these differences were in the predicted direction. In the case of nasals, assimilated sounds were generally more similar to labial nasals rather than to unassimilated coronal nasals. In the case of stops, the measurements for assimilated sounds were, in the majority of cases, more similar to the labial stops although in case of consonant duration, the average value for assimilated oral stops was smaller than the value of the corresponding unassimilated coronal and labial stops. These patterns suggest an overall strong degree of assimilation in the current word-final stop consonants, especially nasal consonants. For further comparing the degree of assimilation between the words ending in nasal and oral stop consonants, multiple mixed effect analyses, using *lmerTest* package in R, were conducted on the data from the four acoustic measurements of raw F2 value, F2 slope, F2 and F1 difference, and closure duration. The fixed effect of the interaction between Consonant Type (nasal stop, oral

stop) and Assimilation (unassimilated, assimilated) and random intercept for item were included in the models. The results of the analyses showed a significant difference between assimilated and unassimilated consonants for all measurements regardless of the type of the consonant (F2:  $\beta = 203.23$ ,  $SE = 42.32$ ,  $t = 4.8$ ,  $p < 0.001$ ; F2 slope:  $\beta = 131.33$ ,  $SE = 29.27$ ,  $t = 4.49$ ,  $p < 0.001$ ; F2 and F1 difference:  $\beta = 228.85$ ,  $SE = 45.57$ ,  $t = 5.02$ ,  $p < 0.001$ ; Duration:  $\beta = 10.17$ ,  $SE = 2.99$ ,  $t = 3.39$ ,  $p < 0.001$ ). The analysis on consonant duration also showed a significant main effect of consonant type ( $\beta = -14.56$ ,  $SE = 4.03$ ,  $t = -3.61$ ,  $p < 0.001$ ) and a significant fixed effect of interaction between Consonant Type and Assimilation ( $\beta = 40.44$ ,  $SE = 4.23$ ,  $t = 9.56$ ,  $p < 0.001$ ). The analyses did not show any other significant effects ( $p > 0.07$ ). A post-hoc least-square mean pairwise comparison (*lsmeans* package in R, V 2.23.5, Lenth, 2016) on the interaction term for consonant duration data showed a significant difference between nasal and oral stops in both unassimilated ( $SE = 4.03$ ,  $p < 0.001$ ) and assimilated conditions ( $SE = 4.03$ ,  $p = 0.002$ ). However, an inspection of the average duration of consonants in each of these conditions suggests that there might be a greater difference in duration of unassimilated coronal and assimilated oral stops ( $M = 199$  vs.  $149$  ms) compared to nasal stops ( $M = 174$  vs.  $163$  ms). To test this observation, the magnitude of the difference between the duration of the unassimilated and assimilated forms of the two types of consonants was calculated for each word. The results of a one-way analysis of variance (ANOVA in R) with the factor Consonant Type (nasal stop, oral stop) on the magnitude of the difference in duration of the consonants showed that this difference is highly significant ( $F(1, 94) = 91.31$ ,  $p < 0.001$ ). The results of the ANOVA tests on the magnitude of the difference based on the other three acoustic measurements did not show any significant difference between nasal and oral stops ( $p > 0.1$ ).

The current results of the comparisons between unassimilated and assimilated coronal stop consonants are in accordance with the results of previous acoustic analyses discussed earlier in the chapter (Gow, 2002; Gow & McMurray, 2007; Dilley & Pitt, 2007). Acoustic measures of F2, namely raw F2 value and F2 slope (F2 transition) as well as the relative measure of F2 and F1 difference, which was only included in the current study, all showed sensitivity to the modifications to the place of articulation of both nasal and oral stop consonants. In addition, consonant duration, which was also reported in Dilley and Pitt (2007), was shown to be a strong indicator of modifications to the canonical form of the oral and nasal stop consonants. The current results however, differ from these previous studies, in that three out of four

measurements did not indicate a large difference between assimilated coronal and unassimilated labial stop consonants in the current stimuli. Even though measurements corresponding to assimilated nasal or oral stops were on average somewhat in-between the measurements for unassimilated labial and coronal consonants (especially for oral stops), the difference was relatively small. The exception to this pattern was consonant duration, which was a strong indicator of the difference between unassimilated labial and assimilated oral stop consonants. This is unlike previous studies, where the measurements of assimilated stops were found to be in-between the measurements of unassimilated labial and coronal stops. Additionally, unlike in the previous studies, where the analyses were conducted on a mixed group of oral and nasal stops, in the current analyses, the two groups were separated and the results were compared. The results of the comparison between nasal and oral stop consonants revealed that in fact the two groups, both containing strongly assimilated forms, substantially differ in terms of at least one of the measured acoustic cues, namely consonant duration. While closure duration is strongly correlated with modifications in oral stops (presumably due to lack of release burst in assimilated forms), for word-final nasal stops, it does not show much variation across unassimilated coronal, assimilated coronal and labial nasals. These results highlight the importance of one of the main questions in the current work, that is, considering the differences in the acoustic properties of assimilated forms, whether seemingly similar variations in nasal and oral stop consonants are treated the same or differently by the perceptual system.

## Chapter 3

### Forced Choice Identification Studies

Previous studies have shown the important role that acoustic cues play in the identification of the place of articulation of consonants. These include cues such as formant transitions from vowels to the next sound, consonant duration and release burst (in stop consonants). In everyday language use, phonological and co-articulatory processes often result in various degrees of change in the key acoustic details of speech sounds. However, slight changes in acoustic characteristics of sounds have shown to impact perception even in tasks that do not direct attention to the speech signal (Dahan et al., 2001). For example, Gow and McMurray (2007)'s analysis of the acoustic details corresponding to assimilated versus unassimilated coronal or labial consonants suggested that the place cues in assimilated sounds were intermediate between canonically coronal and labial place of articulation, and listeners could correctly identify potentially ambiguous words (e.g., *cat<sup>p</sup>*) on 90% of trials providing that the following phonological context was viable for assimilation. On the other hand, as noted in Chapter 1, Dilley and Pitt (2007) found no significant difference in acoustic measurements compared between assimilated and canonically labial consonants in many of the naturally assimilated words in the Buckeye corpus data. In keeping with Dilley and Pitt, the acoustic analysis of the assimilated and unassimilated stimuli in the current work, reported in Chapter 2, indicated a minimal, if any, statistical difference between the measured acoustic cues corresponding to the labial and assimilated coronal nasal or oral stop consonants. In the recognition of spoken words, Gaskell and Snoeren (2008) showed that some naturally assimilated words that can be lexically ambiguous (e.g., *run-rum*) were in fact perceived with a final labial (or velar) stop consonant (e.g., *rum*), if no additional information (such as an appropriate semantic context) has been provided for the listener. As such, the role of the factors involved in recognition of assimilated words, especially in cases where strong assimilation might result in lexical ambiguity, should be investigated in relation to the crucial perceptual influence of the acoustic cues to place of articulation.

The goal of the two experiments reported in this chapter was to assess the role that phonetic cues to place of articulation play in the recognition of assimilated and unassimilated words when phonological context information is absent. The studies that have shown effects of phonological

context in the recognition of assimilated words have mostly examined the viability of the context for place assimilation (e.g., Darcy et al., 2009; Gaskell & Marslen-Wilson, 1996; Mitterer et al., 2006). The comparison was usually made between a condition where the following consonant is a trigger for assimilation (e.g., a labial or a velar consonant in English) and a condition where the following context does not allow for assimilation to take place (e.g., a coronal consonant). The results generally agree that only the appropriate type of phonological context that licenses assimilation helps with recognition of assimilated forms. The results, however, arguably can also be interpreted in a different way, suggesting an unviable phonological context hinders the process of compensation for assimilation, rather than showing a viable phonological context helps with the process. To create the unviable context condition, the naturally produced sounds are usually cross spliced with sounds from a different context (e.g., the word *box* from *cat<sup>p</sup> box* would be replaced by the word *drawing* from *cat drawing*; Gow & McMurray, 2007). Consequently, auditory stimuli created this way would contain mismatching phonetic evidence. However, previous research on the effect of mismatching phonetic cues suggests that such cues interfere with recognition of spoken words (e.g., Marslen-Wilson & Warren, 1994; Marslen-Wilson & Zwitserlood, 1989; McQueen, Norris & Cutler, 1999; McQueen & Viebahn, 2007). Even when assimilated words were naturally produced in an unviable context by intentional mispronunciation of the words in connected speech, the resulting auditory stimuli might sound unnatural to the listeners and could potentially interfere with the listeners' natural perception processes. Thus, the relative difference between the unviable and viable context could reflect "disrupted" vs. "neutral" processing, rather than "facilitative" vs. "neutral" effect of context in processing place assimilation.

Experiments 1 and 2, therefore, were conducted for three main purposes. The first was to function as a pre-test to examine whether the assimilated and unassimilated stimuli ending in nasal or stop consonants in the recorded materials did in fact carry the intended perceptual properties. Even though the acoustic analysis conducted on the current experimental stimuli clearly showed a difference between assimilated and non-assimilated sounds in the expected direction (Chapter 2), it was crucial to test whether the distinctions were also apparent perceptually for actual listeners, and to what extent. Second, the forced choice alternative identification task provides the required baseline for interpreting the results of the following two eye tracking experiments, namely Experiments 3 and 4, where listeners heard the exact same



recordings as in Experiments 1 and 2, only this time, the licensing phonological context was also made available. Finally, the third purpose was to compare patterns of identification of unassimilated and assimilated forms across the two groups of nasal and oral stop consonants. Experiment 1 tested the identification of words ending in nasal consonants and Experiment 2 tested words ending in stop consonants.

In these experiments, the recognition of isolated assimilated and unassimilated words was examined using a forced-choice lexical identification task combined with a priming paradigm. Linguistic phenomena have been shown to trigger priming effects at various levels of representation and language processing (Neely, 1977; Pickering & Ferreira, 2008; Radeau, Morais & Deirt, 1989; Zwitserlood, 1996). In this paradigm, exposure to a stimulus (prime) influences the subsequent processing and recognition of related forms or concepts (target). For example, form-based priming effects such as phonological priming have been obtained in studies of spoken word recognition (Goldinger, Luce, Pisoni & Marcario, 1992; Slowiaczek, Nusbaum & Pisoni, 1987). As described by Goldinger et al. (1997), phonological priming can *inhibit* target recognition when the prime word is similar to the target word in terms of phonetic features but has no shared phoneme with the target (e.g., *bull-veer*, where [b] and [v] share labial features, [u] and [i] are both high vowels, and [l] and [r] are both liquids). Another result can be *facilitation* in recognition when the prime and target words share at least one phoneme (e.g., *bull-bat*, where [b] is shared between the two words). Previous studies using this type of form priming paradigm suggest that the facilitation resulting from overlap in word-final segments (e.g., *bean-seen*) is very short-lived. For example, if there is one intervening item between the two words, those effects were no longer observed (e.g., Slowiaczek, McQueen, Soltano & Lynch, 2000). Furthermore, studies do not find the phonological priming effect in cases where there is word-final overlap between prime and target words in only one segment (e.g., *read-side*; Slowiaczek et al., 2000; Slowiaczek, Nusbaum & Pisoni, 1987). In addition to phone-based priming, studies focusing on syntactic structures and processes have reported *structural* priming effects at higher processing levels (e.g., Pickering & Ferreira, 2008). Structural priming at the level of sentence structure is generally defined as improved performance in processing (e.g., shorter RT) or producing a sentence that is similar in grammatical structure (e.g., passive versus active voice) to another sentence that was previously heard. These structural effects are not limited to higher levels such as syntax, however. Priming of syllabic structures such as word-initial CV structures

are also reported (e.g., Neely, 1977; Meeuwissen, Roelofs & Levelt, 2004; Sevald, Dell & Cole, 1995). Another type of priming is called *process priming*, which according to Janiszewski & Wyer (2014) “occurs when the execution of a process at time one (prime) makes that same process more accessible for use in a subsequent cognitive task.” Perhaps process priming in language processing is best known in relation to processing novel conceptual combinations in noun-noun compounds (e.g., *spear chisel*) and novel metaphorical (as opposed to literal) words or sentences (e.g., “My job is a jail”). In studies of metaphor processing, sentences containing metaphorical target words were read more quickly when a preceding sentence also contained a (semantically unrelated) metaphorical word (prime) compared to when it contained a literal word, suggesting comprehension of the metaphor in the preceding utterance (rather than its form or meaning) facilitated the similar comprehension process required for understanding the target metaphor (Inhoff, Lima & Carroll, 1984; Wisniewski & Love, 1998). Here I draw on this effect to explore the mechanisms involved in processing phonological variation. If the underlying processing mechanisms for compensation for place assimilation and the possible phonological context effects are in fact the same for nasal and oral stop consonants, then it might be possible to observe a similar effect of “process priming” for both groups of consonants. For example, because the recognition of an assimilated word is assumed to involve “undoing” the process of assimilation, hearing and correctly recognizing an assimilated prime word could plausibly facilitate the processing of an assimilated target word heard shortly after the prime. Such process priming would be observed through increasing the consideration for coronal interpretation of the word-final stop consonants and/or faster identification of the target word. On the other hand, the recognition of an unassimilated prime word does not involve the process of compensation for assimilation. Therefore, hearing an unassimilated word is not expected to prime the particular strategies involved in undoing place assimilation and facilitating identification of an assimilated target word. If the priming effects on recognition of unassimilated and assimilated targets are obtained when prime words are unassimilated, the effects should be interpreted as form priming—activating the coronal stop form rather than the process of place assimilation.

In the current chapter, the priming paradigm was incorporated into the two experiments (Experiments 1 and 2) in order to test for potential effects that would be triggered by hearing an isolated word with an assimilated or unassimilated consonant (the prime) before hearing the target word. Critically, the target word was varied such that the final coronal nasal or oral stop

consonant was either produced as unassimilated or it underwent a (strong) coronal-to-labial place assimilation. The identification task should also detect any differences in perception of assimilated and unassimilated nasal and stop consonants. If listeners are sensitive to the acoustic differences reported in Chapter 2, they should be more biased toward choosing words ending in a labial rather than coronal consonant upon hearing the assimilated words (e.g., choosing *cap* when hearing *cat<sup>p</sup>*). Moreover, seeing as the acoustic measures showed place assimilation was in fact stronger in the nasal word-list compared to the stop word-list, identification patterns should be even more biased towards word-final labials in the nasal group. On the other hand, if the acoustic differences were not strong enough to influence listeners' perception of speech sounds, then little or no difference in identification responses should be observed across different conditions.

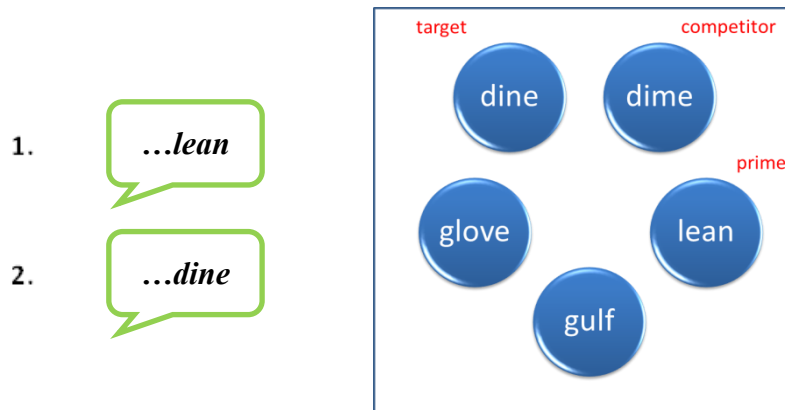
The target stimuli used in the current study consisted of words that could potentially be ambiguous when their final consonant is assimilated. For example, a word like *cat* can potentially become ambiguous, especially when the assimilation is (near) complete, as it would be pronounced similar to *cap*, which is also a word in English. Previous studies on recognition of assimilated words have shown that in fact assimilated words that are potentially ambiguous might be harder to recover and require more processing resources. This is due to the additional lexical competition that needs to be resolved for correct identification of the word (Gaskell & Marslen-Wilson, 2001; Gaskell & Snoeren, 2008; Gow & McMurray, 2007). In the current experiments, response times were measured for both potentially ambiguous and unambiguous words that were correctly identified. This was done to investigate possible differences in processing based on the lexical ambiguity of assimilated words. Response times were expected to be relatively longer for potentially ambiguous words compared to unambiguous words as a result of the confusion that might arise during the lexical access for the former group of words.

In the following sections, I first describe the materials and the general procedure. Then, I present the two experiments and discuss and compare the obtained results within and across the two. In the first experiment, the critical stimuli consisted of (isolated) words ending in assimilated and unassimilated coronal nasals. In the second experiment, the critical stimuli all ended in assimilated and unassimilated oral stop consonants.

### 3.1 General Method and Materials

As noted earlier, the 24 prime and target words had been recorded within two carrier sentences, namely “Click on the \_\_\_ button” and “Now, click on the \_\_\_ button” respectively (see Section 2.1 for more details). For the identification experiments reported in this chapter, the prime and target words were extracted from their carrier sentences. Using Praat software, the words were spliced beginning from the onset of the first consonant in each word to the offset of the nasal murmur (Experiment 1) or at the stop release burst (or stop closure for assimilated forms; Experiment 2). Boundaries were placed at the nearest zero-crossings. Three beep sounds were then added to replace the excised portion of the carrier sentence that originally preceded the test words (1422 ms). The word ‘*button*’ that followed the words in the original recordings was removed. The beep sounds preserved the same 1422 ms interval in the original sentences before the listeners heard the test words. This prepared the listeners for the upcoming stimuli similar to the way in which the original carrier sentences would consistently cue the presentation of a stimulus word.

The visual display for each trial consisted of a set of five orthographically labeled push buttons, presented on a computer screen (Figure 3-1). On critical trials, the five labeled buttons corresponded to five words including a prime (e.g., *lean*), a target (e.g., *dine*), a phonologically-overlapping “competitor” word for the target, and two phonologically and semantically unrelated distractor words. The competitor was the corresponding minimal pair for the target word, always ending in a labial consonant (e.g., *dime*). The relative location of the various button types was fully counterbalanced.



**Figure 3-1.** Example of visual display. The target, competitor, prime and distractor words are represented as labels on five similar buttons. The auditory stimuli consisted of the prime word followed by the target word.

In every trial, two auditory stimuli accompanied the visual display. On critical trials, the first auditory stimulus was the isolated prime word, which always ended in a coronal consonant. The second auditory stimulus (immediately presented after selecting the prime word on the display) was the isolated target word, also ending in a coronal consonant. The competitor for the target word was present in the visual display although it was never referred to auditorily. All prime, target and target competitor words in Experiment 1 ended in a nasal consonant and in Experiment 2 ended in an oral stop consonant. There were 24 critical trials in each experiment. There was a 1000 ms interval between trials from the time when the listener's second selection was made to the automatic initiation of the next trial. The audio and visual stimuli were programmed and presented using Experiment Builder software (SR Research, V 1.10.165B).

The experiments involved a  $2 \times 2$  factorial design. One factor varied the assimilated/unassimilated status of the prime word (in the first auditory stimulus for each display), and the second varied the assimilated/unassimilated status of the target words (in the second auditory stimulus, see Table 3-1). Four lists were created, each containing an equal number of stimulus items in the four conditions. Words were assigned to the distinct lists such that no word was repeated within a list but across the four lists, each word occurred in all four main conditions. In each experiment, an equal number of participants were assigned to the four lists.

**Table 3-1** Example prime and target stimuli from Experiment 1 in the original carrier sentences. The main factors are assimilation in prime words (assimilated-unassimilated) and target words (assimilated-unassimilated).

		Prime carrier sentence	
		<i>unassimilated</i>	<i>assimilated</i>
Target carrier sentence	<i>unassimilated</i>	1. click on the [ <b>lin</b> ] button 2. now click on the [ <b>dajn</b> ] button	1. click on the [ <b>lin<sup>m</sup></b> ] button 2. now click on the [ <b>dajn</b> ] button
	<i>assimilated</i>	1. click on the [ <b>lin</b> ] button 2. now click on the [ <b>dajn<sup>m</sup></b> ] button	1. click on the [ <b>lin<sup>m</sup></b> ] button 2. now click on the [ <b>dajn<sup>m</sup></b> ] button

A secondary factor that was taken into consideration was whether the prime word had a minimal pair competitor. Words being used as primes were assigned such that two out of six words occurring in a given condition could potentially be lexically ambiguous as a result of nasal place assimilation (e.g., *clan-clam*). The minimal pairs of the prime words (e.g., *clam*) were never auditorily or visually present on the display in the critical trials. Another secondary factor that was controlled for in the current experiment was lexical frequency of the prime and target words. The effect of lexical frequency on speech recognition is well known (Meunier & Segui, 1999; Segui et al., 1982) and has often been demonstrated in previous studies using a visual world methodology (Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Tanenhaus, Aslin, & Dahan, 2003). As speech unfolds in time, lexical candidates are activated to a different extent (McClelland & Elman, 1986; Warren & Marslen-Wilson, 1987) and this has not only a relationship with how closely they match the unfolding auditory input but also reflects their frequency of occurrence in speech. In eye tracking studies, it has been observed that high-frequency lexical targets (HF; e.g., *heart*) are initially fixated to more quickly and for a longer duration than low-frequency targets (LF; e.g., *harp*), which is consistent with the idea that HF words are activated more strongly and are recognized more quickly than LF words. Lexical

frequency also has an influence on the probability of occurrence of connected speech alterations. Previous studies indicate that the higher the usage frequency of a word, the higher the probability of it being reduced or modified (Dilley & Pitt, 2007; Scarborough, 2004). Controlling for frequency therefore serves the goal of avoiding artefacts arising from uncontrolled differences in lexical frequency that might affect the perception of assimilated/reduced forms and the effect of priming. To control for lexical frequency, lexical frequency measures were obtained using the SUBTLEXus database (Brysbaert & New, 2009; Appendix D). SUBTLEXus provides a word frequency measure based on 51 million words gathered from American movie subtitles. The advantage, compared to other word frequency indices, is mainly that SUBTLEXus is based on spoken language rather than written corpora and as such, more relevant in analyses of spoken word recognition.

For each critical trial, the frequency difference between the target and competitor word in each trial was calculated and a frequency index was assigned: H-L if the target had a higher frequency than the competitor and L-H if the competitor had a higher frequency. As mentioned above, four subversions/lists of the experiment were created. Within each list, the six target-competitor minimal pairs assigned to a given condition contained three H-L and three L-H pairs. Also, on average, the magnitude of the frequency difference between target-competitor was matched as closely as possible across lists. Prime items were also balanced for their frequency such that, on average, the frequency of the six primes in a given condition was closely similar to those assigned to the other conditions.

Prime words that have a minimal pair ending in /m/, as well as target-competitor word pairs on each critical trial, were also balanced for orthographic complexity such that word pairs in each group like *done-dumb* or *win-whim*, in which one member is more orthographically complex than the other, were spread as evenly as possible across lists and conditions. Each word was given an orthographic complexity index of 0, 1 or 2, with 0 assigned to words whose orthographic form is phonologically transparent (e.g., *sun*), a value of 2 being assigned to words with the greatest degree of mismatch between the written and pronounced forms (e.g., *scene*: three phonemes and five graphemes) and a value of 1 assigned to the words which fall intermediate between these two groups (e.g., *pawn*: three phonemes and four graphemes). These values were used when distributing prime words and target-competitor word pairs to different lists and conditions so that no two pairs with an index difference above zero ended up being

within the same condition group (except in one case where the pairs *scene-seem* and *phone-foam* had to be placed within one group to retain the balance in the lexical frequency across the groups). The stimuli were also controlled for the degree of semantic or phonological relatedness of the prime word to either target or competitor. If a prime has a strong connection with a word that is going to be the target or the target competitor, it could potentially affect the activation level of that word. Therefore, care was taken to avoid using primes that were judged to be semantically or phonologically similar to target-competitor pairs within a given trial.

Because all the critical stimuli ended in either a labial or a coronal nasal or stop consonant, it was possible that the experimental importance of these sounds would become apparent to participants after a few trials. Therefore, filler trials were used to break any expectations and avoid strategic responses. A total of 72 filler trials and two practice trials were added. In twelve filler trials, either the prime, target or both ended in a labial consonant [m] (e.g., *shame*; Exp. 1) or [p] (e.g., *ship*; Exp. 2). This set of filler items served the purpose of neutralizing participants' potential expectation that all critical items end in a coronal consonant. In twelve other trials, (near) minimal pairs, differing in place of articulation of their final consonant (coronal vs. labial), were included in the display but they were never referred to (e.g., *rain-rim*). In another twelve trials, other types of phonologically similar words were displayed but were never referred to (e.g., *heart-harp*). This was unlike the pattern seen in critical trials where the target and competitor items were phonologically similar differing only in place of articulation of their final sound and the target would always be mentioned. In twelve filler trials, phonologically similar targets and competitors shared the coda instead of the onset (e.g., *tack-back*) to draw attention away from the nature of pairs such as *lit-lip* in critical trials that always differed in their coda position. Another twelve trials contained semantically related items (e.g., *pencil, desk*), and in another set of twelve trials, items were orthographically similar/confusing (e.g., *night, sight, might, high*). These filler items were included to distract participants from the focus of the study on sound-level phenomena involving final labial and coronal consonants. Finally, in 30 filler trials, the prime word had a competitor displayed on the computer screen (e.g., *ale-whale*). This was also unlike critical trials, where prime words did not have a phonological competitor in the display. The practice trials were used to familiarize participants with the experimental procedure and included words that were not related to the purpose of the main experiments.



## 3.2 Experiment 1: Nasal Place Assimilation

### 3.2.1 Method

#### 3.2.1.1 Participants

Twenty-four native English speakers were recruited at the University of Toronto. Participants were 20 female and four male adults between 18-40 years of age ( $M = 25$  years), who had no history of hearing or speech difficulty and had normal or corrected vision. Before the experiment, each participant provided a written informed consent and completed a language background questionnaire. Based on the language background data, all participants self-reported native proficiency in English using a 5-point scale, with 5 being the highest level of language proficiency ( $M = 5$ ,  $SD = 0$ ). Participants also reported learning English in early childhood ( $M = 2$  years) and using this language predominantly in their everyday communication. None of the recruited participants were excluded from the final data analysis. All participants received monetary compensation.

#### 3.2.1.2 Materials

The 24 critical trials involved the prime and target words that ended in an either assimilated or unassimilated coronal nasal consonant /n/, along with the list of filler and practice trials (two practice and 72 filler trials). The preparation of the materials and the design for Experiment 1 was the same as described in Section 3.1.

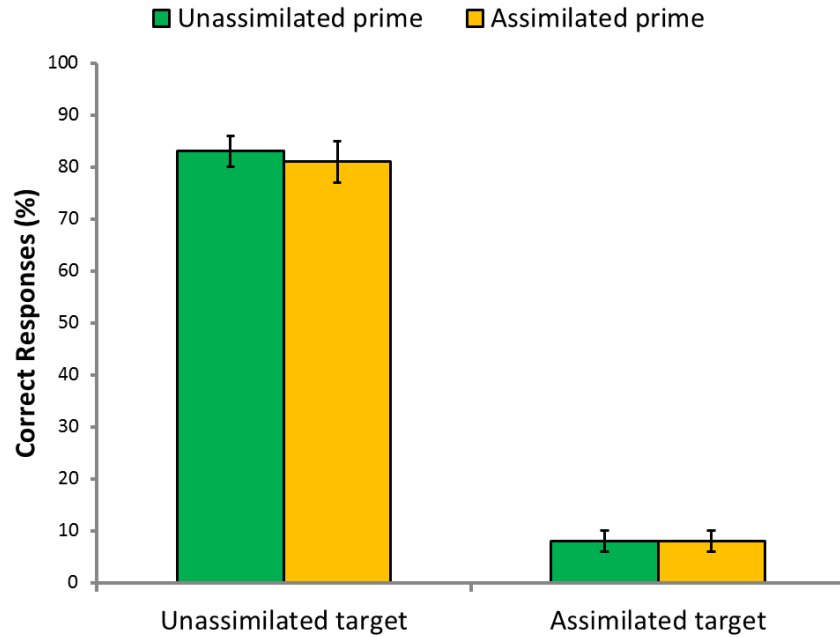
#### 3.2.1.3 Procedure

Each participant was seated individually inside a sound booth in front of a computer screen at a relatively fixed distance (~109 cm) from two loudspeakers. The loudspeakers were positioned from the participant's seat and were calibrated such that the intensity of speech spectrum noise was around 71 dB SPL. Written instructions for the experiment were provided on the computer screen prior to the start of the practice trials. Participants' task was to click with the computer mouse on the displayed labeled button that best matched with the word they would hear after the three beeps preceding each test word. On each trial, the labelled buttons appeared 3000 ms before the first auditory stimulus (prime) played to allow for enough time for scanning the screen. The appearance of a small triangular mouse cursor in the middle of the screen and an accompanying tone then signaled the start of the trial. However, mouse clicks were not effective

until after the onset of the word. After the participant's click response to the prime stimulus, the second auditory stimulus (target) played automatically for the same display. The trial ended after the participant's second selection (target click response) and the next trial immediately followed. The experiment took approximately 20 minutes in total.

### 3.2.2 Results

The average percent of correct responses was calculated for the identification data for both prime and target stimuli. A response was considered correct if the intended prime or target word, which always ended in a coronal nasal, was selected by a participant. The results of the analysis showed that participants correctly recognized the intended prime word on 100% of trials in both the unassimilated and assimilated conditions. This was expected, as no phonologically and/or orthographically similar word to the prime was present on the visual displays. For the target words, however, mean accuracy in the unassimilated condition was 82% (see Figure 3-2). This is due to the presence of the phonological competitor, which only differed from the target word in the place of articulation of the word-final nasal. In the condition where the target was assimilated, participants almost always selected the target competitor word ending in the labial consonant (only 8.3% correct responses), which suggests further misperception resulted from the effect of place assimilation on the final coronal nasals.

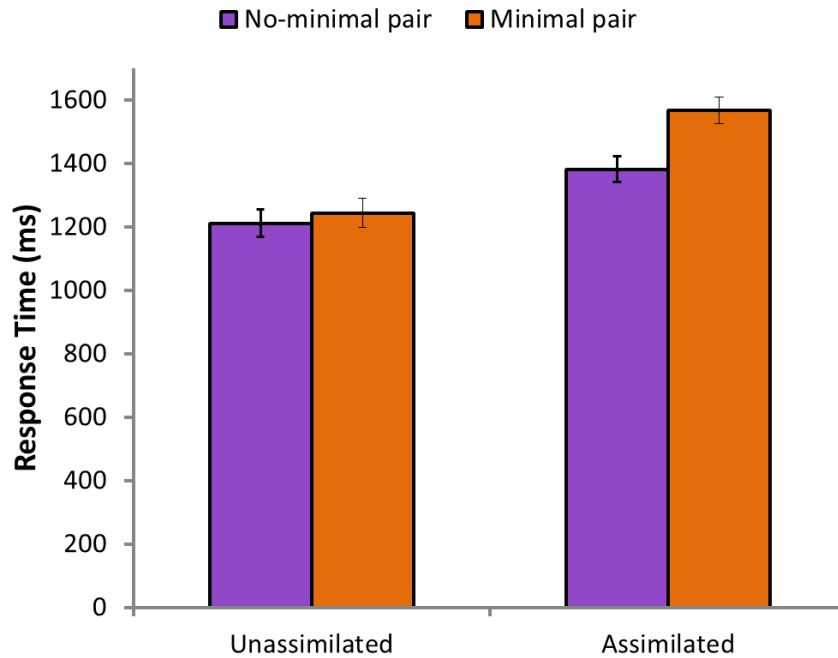


**Figure 3-2.** Percentage of correct responses for isolated target words ending in nasals (Exp. 1), across conditions. The error bars represent the standard error.

A logistic mixed effects model using the *glmer* function and *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016) in R with Prime (unassimilated vs. assimilated) and Target (unassimilated vs. assimilated) and their interaction as fixed effects was conducted on the target word correct response data. Random intercepts for participant and item and a random slope for Target by participant were included.<sup>10</sup> The results showed that assimilated targets were significantly less often identified as ending in a coronal nasal compared to unassimilated targets ( $\beta = 4.25$ ,  $SE = 0.47$ ,  $z = 9.05$ ,  $p < 0.001$ ). The percentage of correct responses did not significantly differ based on the Prime ( $p = 0.9$ ). Also, the interaction of Prime and Target was not significant ( $p = 0.8$ ).

---

<sup>10</sup> Random effects in models were kept maximal as far as possible (Barr, Levy, Scheepers & Tily, 2013). Random intercepts and/or random slopes that resulted in the models failing to converge were excluded from the statistical models reported throughout this work.



**Figure 3-3.** Average response RT for isolated prime words ending in nasals (Exp. 1). Plotted based on assimilation condition for primes with or without a minimal pair. The error bars represent the standard error.

In addition to the identification scores, participants' response time (RT) to click on the intended word was also calculated on correct trials for the prime words only<sup>11</sup>. The RTs were measured from the offset of the word up to the time point when the button corresponding to the prime word was clicked on by the participant. The RTs that were identified as being too slow or too fast ( $\pm 2$  SD from the mean) were marked as outliers and were replaced by the sample average (4.2% of the data; Tabachnick & Fidell, 2007). RTs were calculated separately for unassimilated and assimilated prime words and were further compared based on whether assimilation could result in lexical ambiguity (Fig. 3-3). Recall that out of 24 prime words, eight words had a

---

<sup>11</sup> Note that RTs for target words were not calculated in the current paradigm. This is because to click on the button for the target word, participants needed to move the cursor from where it was resting after they clicked on the preceding prime button. Because the relative position of prime and target buttons was counterbalanced, the mouse travel distance between these buttons varied considerably across trials. In contrast, with the prime words, the cursor was always in the middle of the screen at the time the word was heard; therefore, the mouse travel distance did not vary across trials.

phonological minimal pair (differing only in the word-final consonant) and could potentially become lexically ambiguous as a result of place assimilation (e.g., *bean* can sound similar to *beam* in “bean button”). A linear mixed effects analysis using *lmer* function from *lmerTest* (Kuznetsova et al., 2015) in R, with Assimilation (unassimilated vs. assimilated), Minimal Pair (minimal pair vs. no-minimal pair) and their interaction as fixed effects, was conducted on the RT data for correct trials (100% of the trials in the current experiment). Random intercepts for both participant and item were included, as well as random slopes for Assimilation and Minimal Pair by participants. The results indicate slower RTs in the assimilated condition ( $M = 1474$  ms) compared to the unassimilated condition ( $M = 1228$  ms;  $\beta = -322.15$ ,  $SE = 46.11$ ,  $t = -6.99$ ,  $p < 0.001$ ). Also, the RT for words with a minimal pair was on average slower ( $M = 1405$  ms) than for no-minimal pair words ( $M = 1297$  ms;  $\beta = -184.08$ ,  $SE = 51.23$ ,  $t = -3.59$ ,  $p < 0.001$ ). There was also a significant interaction effect ( $\beta = 150.72$ ,  $SE = 56.03$ ,  $t = 2.69$ ,  $p = 0.007$ ). Post-hoc least square means pairwise comparisons (*lsmeans*, Lenth, 2016) were conducted on the significant interaction term. The results showed that on average RTs for primes with a minimal pair ( $M = 1566$  ms) were significantly slower than the RTs for primes with no minimal pair ( $M = 1382$  ms) when the words were assimilated ( $SE = 184.08$ ,  $p = 0.005$ ) but not when they were unassimilated ( $SE = 33.36$ ,  $p = 0.91$ ).

In sum, the higher rates of misidentification of the assimilated target stimuli as ending in labial rather than the underlying coronal nasal suggest that the degree of perceived assimilation in the current stimuli was strong/complete. This matches with the results of the acoustic analysis conducted on the current set of stimuli (refer to Section 2.2.1.1), which showed a relatively strong degree of place assimilation in the words ending in nasal consonants. In addition, in the presence of a phonological competitor, words were less often correctly identified even when they were in their canonical form. This further suggests that in some cases the acoustic cues to place of articulation of word-final nasals might not be enough for disambiguation of the isolated minimal pairs that differ only in the place of articulation of their final nasal consonant (refer to Section 1.1.1.1 for a review on acoustic and perceptual characteristics of nasal stops). Moreover, identification scores suggest that there does not seem to be any effect of having exposure to either an unassimilated or assimilated prime on recognition of the following assimilated target words. Finally, slower RTs when the prime words were assimilated, especially when there was a potential for lexical ambiguity, suggests a higher processing load as a result of an extra step that

supposedly should be taken by the listeners to “undo” the phonological process of assimilation and to overcome the created lexical competition to access the lexical item. This is despite the fact that there was no competitor present in the display for the prime words.

The goal of Experiment 1 was to investigate the perceptibility of naturally produced unassimilated and assimilated word final nasal consonants when words were presented in isolation. In this case, the only source of information for listeners to recover the underlying place of articulation of the final nasal would be the residual acoustic cues of the sound itself. However, differences in the saliency of the acoustic cues to place of articulation as a function of consonant type (nasal vs. oral stop), as well as the degree of assimilation, might affect the perception of final consonants. Experiment 2, therefore, investigates the same question regarding the perceptibility of the unassimilated and assimilated word final consonants focusing on oral stops, and how this sound group might differ from nasal stops.

### 3.3 Experiment 2: Stop Place Assimilation

#### 3.3.1 Method

##### 3.3.1.1 Participants

Twenty-four native English speakers, who did not participate in Experiment 1, were recruited at the University of Toronto. Participants were 16 female and eight male adults between 18-35 years of age ( $M = 21$  years) with no history of hearing or speech difficulty and with normal or corrected vision. Each participant provided written informed consent form and completed a language background questionnaire prior to the experiment. All participants included in the final data analysis reported high native proficiency in English on a scale of 1-5, with 5 being the highest level of language proficiency ( $M = 4.88$ ,  $SD = 0.33$ ). Participants also reported learning English in early childhood ( $M = 0.8$  years) and using this language predominantly in their everyday communication. Based on these criteria, nine additional participants were excluded from the final data analysis as they did not begin learning Canadian English until later childhood. All participants received monetary compensation.

##### 3.3.1.2 Materials

The auditory and visual materials that were used in Experiment 2 were prepared the same way as the materials used in Experiment 1, described in Section 3.1. The auditory stimuli consisted of

the list of 24 prime and 24 target words that ended in an either assimilated or unassimilated coronal stop consonant /t/, along with the two practice and 72 filler items. The visual display was similar to Experiment 1 except that the competitor for the target word always ended in a labial stop consonant /p/ (e.g., *cap*).

### 3.3.1.3 Procedure

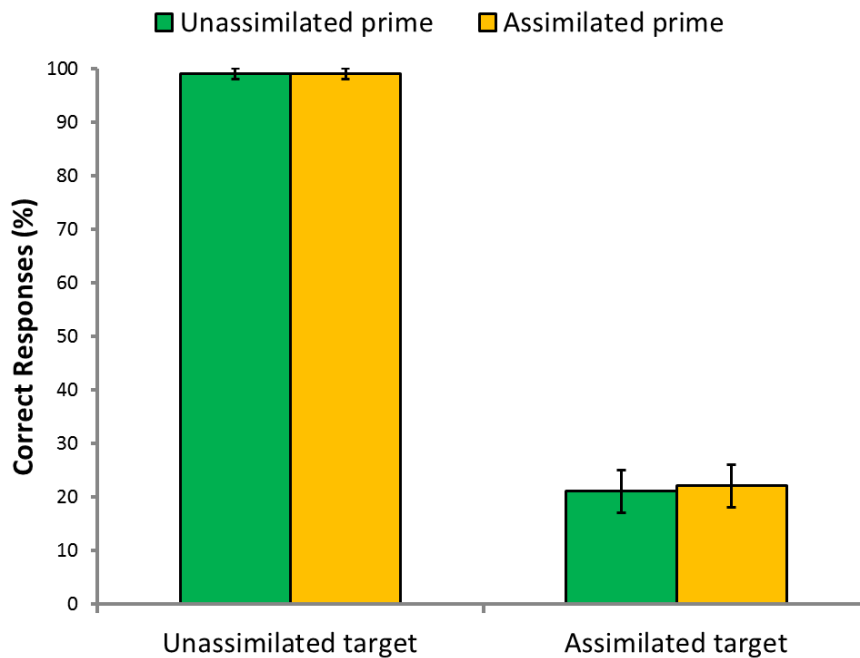
The procedure was the same as in Experiment 1. Participants were individually tested in a sound attenuated booth. Written instructions were provided on the computer screen prior to the practice trials. Participants then completed the trials at their own pace. Each experimental session took approximately 20 minutes in total.

### 3.3.2 Results

Similar to Experiment 1, the percentage of correct responses for identification of prime and target words was calculated in each of the experimental conditions for each participant. Thus, if the participant selected the intended prime or target word ending in a coronal stop, the response was taken as correct. In addition, participants' average response times (RT) were also calculated for unassimilated and assimilated prime words.

Figure 3-4 shows that, on average, participants correctly recognized the intended assimilated or unassimilated isolated prime words 99.6% of the time. This near-ceiling performance for primes is expected due to the lack of a phonologically and/or orthographically similar word to prime words in the set of displayed buttons. For unassimilated target words, identification scores were also close to ceiling (99% correct) in spite of the presence of a minimal-pair competitor in the display. However, when the target was assimilated, the rate of the correct responses dropped to only 21% on average, which suggests place assimilation led the majority of the listeners to rely on the surface acoustics as the primary basis for their judgments. This resulted in the misidentification of the words as ending in a labial stop instead. A logistic mixed effects model (*glmer* in *lmerTest* package using R) was conducted on the correct response data for words ending in stop consonants, with Prime (unassimilated vs. assimilated) and Target (unassimilated vs. assimilated), and their interaction as fixed effects and random intercepts for participant and item and a random slope for Target by participant. The results indicated that the difference that was observed between percentage of correct responses to assimilated and unassimilated target

words was significant ( $\beta = 9.41$ ,  $SE = 3.76$ ,  $z = 2.5$ ,  $p = 0.01$ ). No significant main effect of Prime ( $p = 0.9$ ) or an interaction was found ( $p = 0.6$ ).

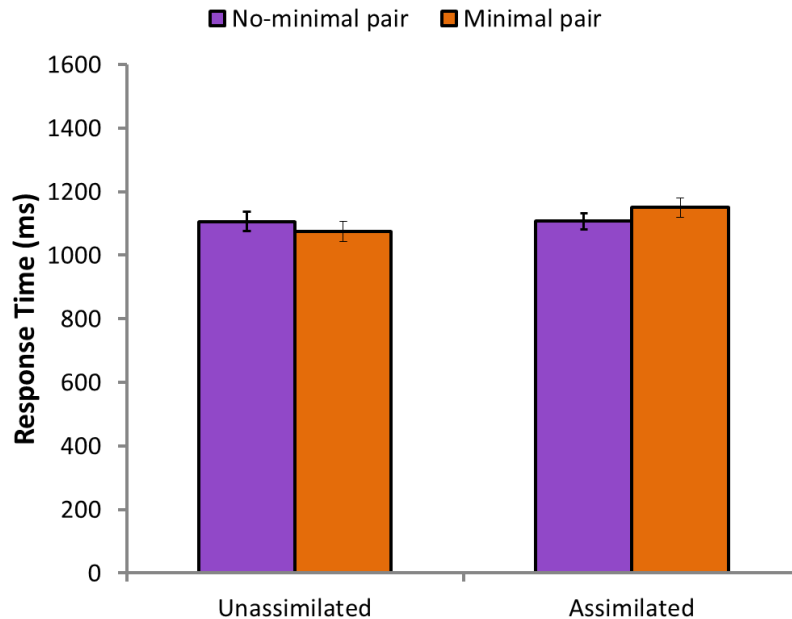


**Figure 3-4.** Percentage of correct responses for isolated target words ending in oral stops (Exp. 2). Plotted based on the assimilation condition for prime and target words. The error bars represent the standard error.

Participants' average mouse click response times for correctly identified prime words were calculated after replacing outlier data points (i.e., RTs shorter/longer than  $\pm 2$  SD above the mean) with the sample average for prime RTs (affecting only 2.3% of all observations). RTs were compared for unassimilated and assimilated versions of the prime word as well as for prime words with and without a minimal pair ending in /p/ (undisplayed; e.g., *rat* versus *fat* respectively, see Figure 3-5). RTs were on average numerically faster for the unassimilated prime words than the assimilated primes (25 ms faster), but the magnitude of this effects is obviously quite small. A mixed effects model (*lmer* from *lmerTest* package in R), with Assimilation (unassimilated, assimilated), Minimal Pair (no-minimal pair, minimal pair) and their interaction as fixed effects was conducted on the RT data. Random intercepts for both participant and item and random slope for Assimilation and Minimal Pair by participant were included in the analysis. The results indicated marginally significant longer RTs in assimilated



condition ( $M = 967$  ms) compared to unassimilated condition ( $M = 827$  ms,  $\beta = -73.88$ ,  $SE = 37.45$ ,  $t = -1.97$ ,  $p = 0.05$ ). There was no significant main effect of Minimal Pair ( $p = 0.2$ ) or a significant interaction ( $p = 0.1$ ).



**Figure 3-5.** Average correct response RT for isolated prime words ending in oral stops (Exp. 2). Plotted based on assimilation condition for primes with or without a minimal pair.

The low percentage of correct target selections for words that end in assimilated stops (21%) is an indicator of a strong degree of assimilation in the current stimuli. When heard without the following phonological context in the original recordings, the word-final assimilated stops tended to be identified as having a labial place of articulation. This outcome is interesting in that the acoustic analyses conducted on the same set of stimuli (reported in Section 2.2.1.2) showed that the assimilated stops might carry certain acoustic cues that are between labial and coronal places of articulation (e.g., consonant duration). However, listeners tend to perceive these sounds as closer to labial than to coronal place of articulation. Also, identification scores suggested that being pre-exposed to a word-final oral stop with ambiguous cues to place of articulation (i.e., hearing the assimilated prime word) did not seem to have any significant effect on the recognition of the following assimilated word. Finally, the prime RT results showed that the processing time is longer for words ending in assimilated stop consonants compared to words

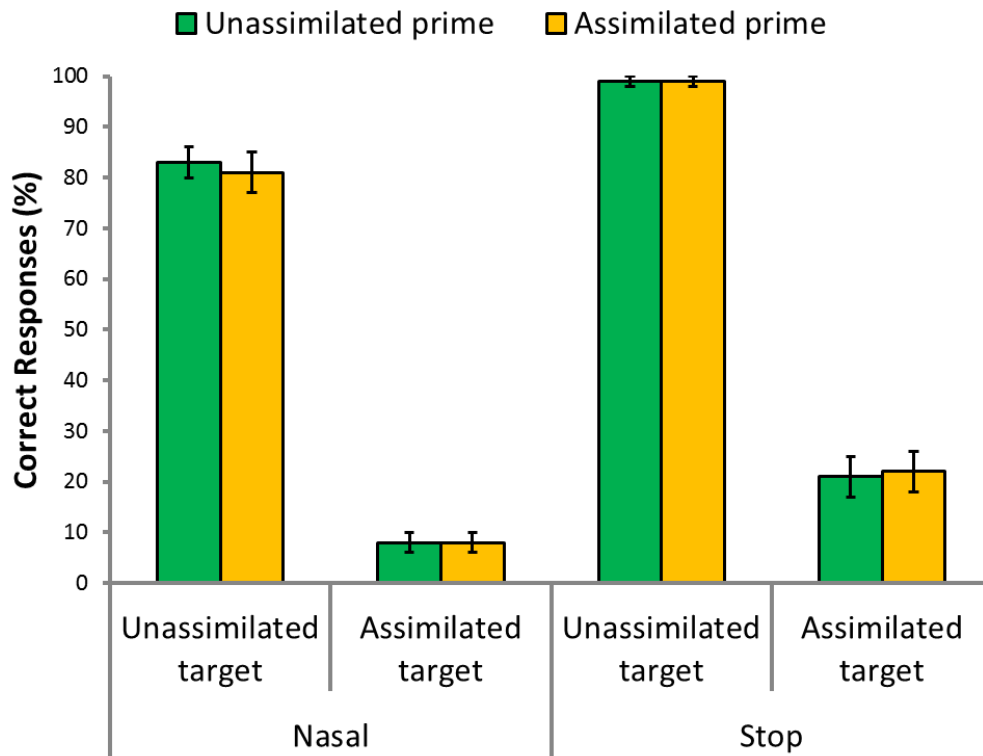
ending in unassimilated stop consonants. Mean RTs were not significantly affected by whether the prime had a minimal pair or not.

### 3.4 Cross-Experiment Comparisons and General Discussion

The aim of Experiments 1 and 2 was to investigate the recognition of words ending in either assimilated or unassimilated nasal or stop consonants when the triggering phonological context for assimilation was not present. A core question was whether and how listeners are able to identify a target word containing a canonical or assimilated word-final nasal or oral stop based on acoustic cues to place of articulation. To this end, both offline identification scores and response times (for prime sentences) were analyzed. Together with the results of the acoustic cue analysis, these results provide a clear understanding of the degree of assimilation of word-final nasal and oral stops in the current stimuli, as well as similarities and difference in their processing. In addition, the results will be used to establish a baseline for further comparison in the remaining experiments, where the relevant phonological context is provided.

The results of Experiment 1 suggest that the assimilated word-final nasal consonants in the current set of stimuli have undergone (near-)complete place assimilation. When assimilated target words were presented without the triggering phonological context, they were almost always misidentified as their corresponding competitor word ending in a labial nasal. This suggests that any residual acoustic cues to coronal place of articulation that were potentially carried by the speech signal were not salient enough to be picked up by listeners when identifying the test word. The results of Experiment 2 also suggested a strong degree of assimilation in word-final stop consonants (e.g., *cat*), although when the results of the two experiments are compared, listeners are slightly more accurate in identifying the underlying place of articulation for stops compared to nasals (Figure 3-6). A logistic mixed effects model (*glmer* in R) was conducted on the pooled target accuracy measures from both nasal and oral stop groups. The fixed factors were Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated), Consonant Type (nasal vs. stop) and their interactions. Random intercepts for participant and item and random slopes for Target and Prime by participant and by item were also included. The results demonstrated a significantly higher number of coronal click responses in the unassimilated target condition, as was expected ( $\beta = 4.61$ ,  $SE = 0.62$ ,  $z = 7.43$ ,  $p < 0.001$ ). Also, the results showed that listeners, regardless of assimilation condition, correctly identified

the underlying coronal place of articulation for word-final oral stops more often than for word-final nasal consonants ( $\beta = 1.18$ ,  $SE = 0.46$ ,  $z = 2.54$ ,  $p = 0.01$ ). None of the other main or interaction effects were significant (all  $p$ 's  $> 0.07$ ).



**Figure 3-6.** Percentage of correct responses for isolated target words ending in nasals (left panel, Exp.1) or oral stops (right panel, Exp. 2). Plotted based on the assimilation condition for prime and target words. The error bars represent the standard error.

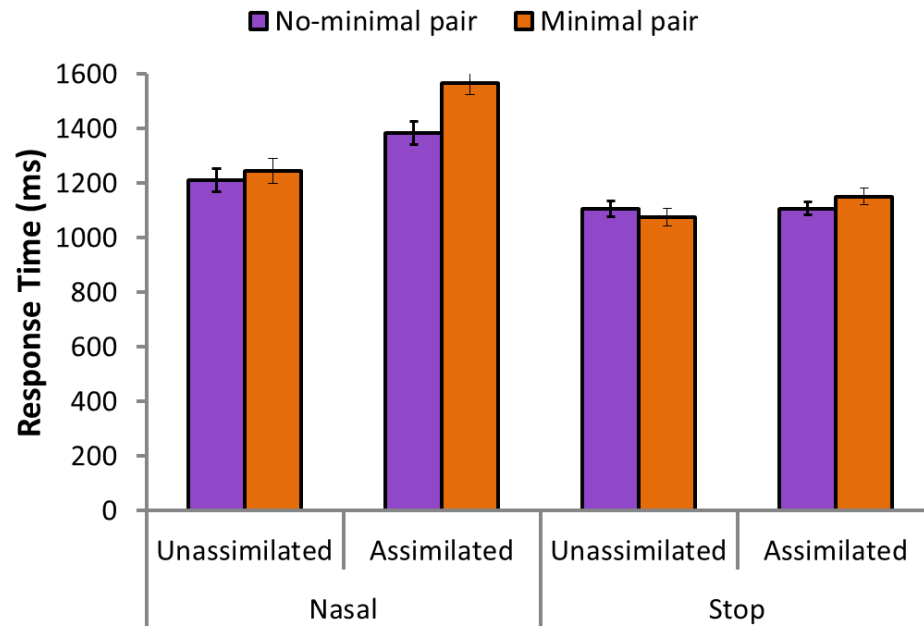
The patterns observed in these two experiments match the pattern of results found in the acoustic analyses conducted on the same sets of stimuli (Chapter 2). Specifically, these analyses showed that the acoustic cues to place of articulation in assimilated nasals were highly similar to the cues that characterize a labial place of articulation. This would explain the high rate of identification of assimilated nasals as labials. In assimilated stops, however, the cues were somewhat intermediate between the corresponding cues to a coronal versus labial place of articulation, potentially providing cues to the underlying place of articulation of the stimulus. The patterns of

acoustic measures reported for stop consonants in the current study are in fact more similar to patterns reported for the stimuli used in previous studies of assimilated consonants (Gow & McMurray, 2007).

Interestingly, comparing the results from identification of potentially ambiguous words ending in nasals with the words ending in oral stops reveals an overall lower perceptibility of nasals compared to stops, even in the *unassimilated* condition. This pattern is in line with evidence from both perception studies and acoustic analyses showing that the cues to place of articulation for oral stops are perceptually more salient than those for nasals (Nolan and Kerswill, 1992; Malecot, 1956; Winters, 2002). In other words, labial and coronal nasals are inherently more perceptually confusable than labial and coronal oral stops.

Additional evidence supporting the perceptual differences between nasal and oral stop consonants comes from a comparison of the response times for unassimilated and assimilated prime words, particularly when there is potential for lexical ambiguity (Figure 3-7). The results from both experiments show that RTs were overall slower for the words ending in nasals ( $M = 1351$  ms) compared to the words ending in stop consonants ( $M = 1110$  ms). A mixed effects model using *lmer* from *lmerTest* package in R, with Assimilation (unassimilated vs. assimilated), Minimal Pair (minimal pair vs. no-minimal pair), Consonant Type (nasal vs. stop) and their interactions as fixed effects was conducted on the RT data from Experiments 1 and 2. Random intercepts for both participant and item were included, as well as random slopes for Assimilation and Minimal Pair by participant. The results show that the difference between nasal and oral stops was in fact significant ( $\beta = -416.24$ ,  $SE = 59.65$ ,  $t = -6.98$ ,  $p < 0.001$ ). There was also a significant interaction effect between Consonant Type and Minimal Pair ( $\beta = 140.91$ ,  $SE = 60.24$ ,  $t = 2.33$ ,  $p = 0.02$ ), between Assimilation and Minimal Pair ( $\beta = 150.72$ ,  $SE = 51.08$ ,  $t = 2.95$ ,  $p = 0.003$ ), and between Consonant Type and Assimilation ( $\beta = 248.35$ ,  $SE = 59.49$ ,  $t = 4.18$ ,  $p < 0.001$ ). No significant interaction was found between the three factors ( $p = 0.3$ ). A series of post-hoc comparisons using least square means (*lsmeans* in R) that were conducted on the interaction effects showed that RTs were slightly slower when they heard words ending in unassimilated nasal consonants compared to words ending in unassimilated stop consonants regardless of the Minimal Pair condition ( $SE = 51.52$ ,  $p = 0.05$ ). Also, RTs to assimilated nasals were slower than RTs to assimilated oral stops ( $SE = 47.17$ ,  $p < 0.001$ ). Furthermore, RTs were found to be slower for prime words ending in nasals compared to words ending in oral stops

whether the prime had a minimal pair ( $SE = 53.78, p < 0.001$ ) or did not have a minimal pair ( $SE = 49.71, p = 0.001$ ). These results suggest that perception of the place of articulation for nasal consonants generally requires a longer processing time and this is especially the case when the consonant is assimilated.



**Figure 3-7.** Average correct response RT for isolated prime words ending in nasals (left panel) or stops (right panel). Plotted based on assimilation condition for primes with or without a minimal pair.

The pattern in the RT results might be an indicator of a difference in perceptibility of the two sounds where the underlying place of articulation of a word-final assimilated stop is generally easier to perceive than its nasal counterpart, hence overall faster RTs in the former case. A slower RT could also indicate a higher level of complexity of the task, which in the current experiments, can be associated with an extra processing step involved in “undoing” assimilation in word-final nasals compared to oral stops. As discussed in Chapter 2, in the current study, assimilated coronal nasals are acoustically very similar to the labial nasal stops, potentially making the recognition of minimal pairs ending in coronal or labial nasal stops more challenging. In addition, as noted in earlier sections, unlike for words ending in nasals, RTs did

not significantly differ based on whether the word had a minimal pair competitor or not for assimilated prime words ending in stops. Combined with the overall slower RT for words ending in assimilated nasals, this outcome suggests there is more lexical competition in the nasal case compared to words ending in assimilated oral stops. Alternatively, the faster RT that was observed for primes ending in oral stops compared to the primes ending in nasals could possibly be explained by the possibility of a difference in the lexical frequency of the two groups. To examine this possibility, the average lexical frequency of the prime words (obtained from SUBTLEXus database; Brysbaert & New, 2009) was obtained for words ending in nasals ( $M = 20.8$ ,  $SD = 32.1$ ) and oral stops ( $M = 37$ ,  $SD = 53.6$ ). A two-sample (Welch's) t-test on the frequency measures, however, indicated no significant difference between the groups in the average lexical frequency ( $t(44) = 0.21$ ,  $p = 0.15$ ). This shows that the difference that was observed in RT results was not likely to be due to the differences in the word usage frequencies.

The results of Experiments 1 and 2 provided a baseline for interpreting the results of the two experiments to be discussed in the following chapter. As mentioned earlier, the focus of Experiments 3 and 4 is to explore how real-time word recognition takes place in a scenario where the following phonological context (a following word with an initial labial consonant) is provided. In these experiments, an eye tracking methodology is used in combination with a similar identification task used in Experiments 1 and 2 to provide time-course information for the processing of assimilated and unassimilated words as speech unfolds in real time.

## Chapter 4

### Visual World Eye Tracking Experiments

As discussed in Chapter 1, previous research has provided evidence that viable phonological context is used in recognition of words containing assimilated consonants. This evidence tends to come from studies demonstrating the contrast in listeners' performance when the same assimilated consonant occurs in a context that should not trigger assimilation (mismatching contextual information; although see Snoeren, Segui and Halle, 2008 for an alternative approach). More specifically, these studies show that, when the phonological context for place assimilation is not viable (as a result of cross splicing the assimilated word between viable and unviable phonological contexts or an intentional (mis)pronunciation), listeners are less likely to perceive an assimilated sound as a coronal rather than a non-coronal sound (i.e., *cat<sup>p</sup> drawing* is perceived as *cap drawing*). In contrast, when the context is viable for assimilation (e.g., the following sound is a labial/velar consonant), listeners are more likely to perceive the same assimilated sound as coronal (i.e., '*cat<sup>p</sup> box*' perceived as '*cat<sub>t</sub> box*'). This indicates that mismatching phonological context has a negative effect on accessing the underlying form of assimilated sounds. This in turn is taken as evidence in favor of the important role that the phonological context information plays in processing assimilation in general. These studies do not, however, provide a direct answer to the question of whether a viable phonological context in fact has a positive influence on compensation for assimilation. That is, because assimilation does not naturally occur in a context that does not trigger assimilation, it is possible that the viability effect shown in these experiments arises as a result of a perceptual "confusion" due to mismatching information in the unviable condition. The current experiments (Experiments 3 & 4) report the case of recognition of assimilated words followed by a viable post-assimilation context. The results are then interpreted in light of a series of comparisons made against the baselines provided by Experiments 1 and 2, where the post-assimilation context was not present in the recordings heard by listeners. Any positive effects of phonological context on recognition of assimilated words can be understood as more straightforward evidence of a genuine regressive contextual effect.

As noted earlier, one important question with regard to the effect of the phonological context on processing assimilated forms is about the precise mechanism that underlies the integration of

contextual information. If general auditory processing mechanisms are primarily or exclusively involved in compensating for place assimilation, the prediction would be that all similar types of variation must be processed similarly, regardless of the statistical distribution of sounds and occurrence of phonological processes within a specific language system. On the other hand, explanations involving top-down integration of probabilistic phonological knowledge would predict contextual effects that are modulated by specific sound properties and language systems (e.g., Gaskell, 2003; Mitterer, Kim & Cho, 2013). Previous studies have mainly focused on cross-linguistic comparisons to search for evidence in favor of language dependent or language independent mechanisms for compensation for place assimilation (Darcy, Ramus, Christophe, Kinzler & Dupoux, 2009; Clayards, Niebuhr & Gaskell, 2015; Mitterer, Csepe & Blomert, 2006). However, the inevitable differences across languages in terms of their phonological systems and the acoustic and distributional properties of speech sounds can make comparisons and interpretations of results less than straightforward. For example, the stimuli used in previous experiments often contain non-native phonemes/distinctive features or non-words with or without intentional mispronunciation of the sounds to produce the effect of a particular phonological process (e.g., Mitterer, Csepe & Blomert, 2006; Darcy, Ramus, Christophe, Kinzler & Dupoux, 2009). A complementary approach is to examine the mechanisms involved in processing variation *within* a language by focusing on whether the effects of contextual information on compensation for a phonological process such as place assimilation differ for different groups of sounds. One focus of the current investigation is therefore to explore processing mechanisms involved in perception of two separate groups of sounds, namely nasal and oral stop consonants, that show distinct properties and distributions in the English language system, but which are both traditionally described in terms of undergoing the process of place assimilation.

In the current research, this question was explored using a visual world methodology. The visual world eye-tracking paradigm has shown to be especially suited for studies focusing on the course of activation of lexical items as speech unfolds (see Tanenhaus, Spivey, Eberhard & Sedivy, 1995). In the current experiments, this methodology was used to track the use of fine-grained acoustic cues and the integration of contextual information during the recognition of assimilated words. Of interest are participants' patterns of eye movements at specific points and/or periods of time when acoustic cues to place of articulation (such as vowel transition cues to the final



nasal or stop consonants), as well as the cues for the upcoming consonant (i.e., the phonological context for place assimilation) become available. Further, in the current experiments, the visual world methodology was combined with a priming paradigm. Here, the eye-movement measures can be used to assess whether processing an assimilated form would be in any way affected by a recent experience requiring compensation for assimilation (i.e., process priming), regardless of the listeners' conscious final decision. When an assimilated target word follows an assimilated prime word (matching prime), facilitatory priming effects might result in an increase in the number of fixations to the target word and/or earlier fixations as the assimilation context becomes available. Since in the current experiment, the prime and target words are presented in separate utterances (i.e., several words intervening the two words) and also there is no phonological overlap between the prime and target words except for the word-final phoneme, no form priming (i.e., phonological priming) effect would be expected. Even if an overlap between prime and target words in their word-final phoneme could trigger phonological priming effects, the effects should occur when the prime word ends in an unassimilated (canonically coronal) consonant and as soon as the acoustic cues to the word-final segment become available, earlier than the acoustic cues to the following context are heard. This is because, as discussed earlier, eye movements are shown to be time-locked to perceptual details as speech unfolds. Therefore, any later effects of priming are expected to be related to the following context and the processing of assimilation.

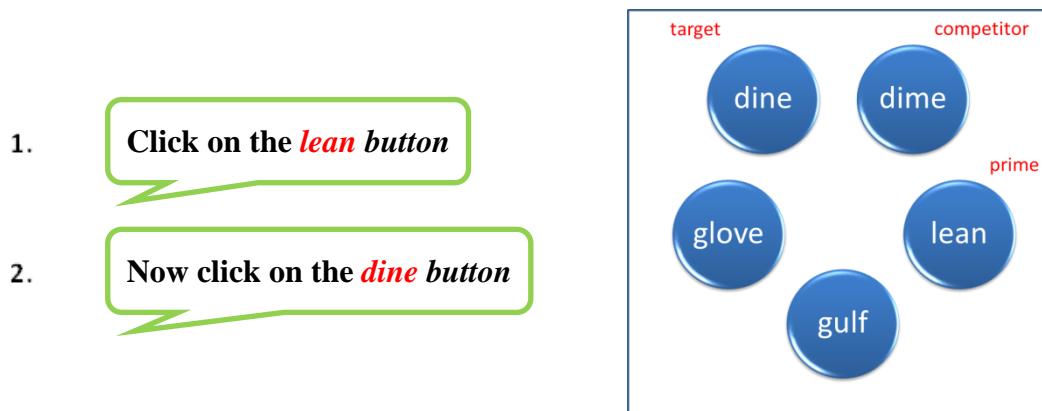
In Experiments 3 and 4, listeners' eye movements were monitored as they heard two sentences carrying the prime and target lexical items. Listeners' task was to choose one of the five labeled buttons on a computer screen that best matched each auditory stimulus. The stimuli in Experiments 3 and 4 contained the same test words that were used in Experiments 1 and 2 respectively. However, in the current experiments, the test words were presented within their original carrier sentence and phonological context (e.g., "Click on the *cat*' button"). The focus of Experiment 3 was nasal place assimilation and stop place assimilation was the focus of Experiment 4.

## 4.1 General Method and Materials

To compare results across the full set of experiments, the general methodology applied to Experiments 3 and 4 was similar to Experiments 1 and 2. The same list of prime-target word

pairs (see details in Section 2.1) used in the first two experiments occurred in the current set of experiments as well. All prime and target words ended in an either unassimilated or assimilated coronal nasal /n/ (Experiment 3) or an unassimilated or assimilated voiceless coronal stop /t/ (Experiment 4). Unlike Experiments 1 and 2, where the prime and target words were spliced out of the original carrier sentences and presented to the participants in isolation, in Experiments 3 and 4, the prime and target words were presented within their original carrier sentences. The carrier sentence for the prime words was “Click on the \_\_\_ button” and for the target words was “Now click on the \_\_\_ button”, where primes and targets replaced the blanks. The sentences were acoustically normalized using Praat software (Version 5.3.23) so that the average intensity of each sentence was 70 dB SPL. Both unassimilated and assimilated prime and target words were included to create four experimental conditions (a  $2 \times 2$  factorial design, see Section 3.1).

The experiments were implemented using Experiment Builder software (SR Research, Version 1.10.165B). The visual display used in the eye tracking experiments was identical to the display used in the identification experiments in Experiments 1 and 2. A set of five buttons labelled with the prime, target, competitor and two phonologically and semantically unrelated words were presented on a computer screen, each of which could be virtually "pressed" by the participant (using a mouse) in response to spoken instructions (Figure 4-1). Reflecting the format used in Experiments 1 and 2, a five-picture display, as opposed to a four-picture display (more commonly used in eye tracking studies), was used on each trial. The reasoning is that, after the button associated with the prime in the first instruction is chosen, participants will still be left with four options for the second instruction, making this crucial phase of the trial similar to the scenario used in most other eye tracking studies. As noted, similar to the identification experiments (Experiments 1 & 2), the visual stimuli were presented using an orthographic representation (labelled buttons). Although in most eye tracking studies the visual stimuli are clip-art images, recent studies have shown that orthographic stimuli can be used to examine online spoken word recognition processes (Huetting & McQueen, 2007; McQueen & Viebahn, 2007; Salverda & Tanenhaus, 2010). The use of orthographic forms has the advantage of providing more freedom because one is not limited to using words associated with depictable objects. For example, because the forms are always being represented as labels on a picture of a button, words from different categories like nouns, adjectives, and verbs can be used (e.g., Click on the *gain* button).



**Figure 4-1.** Example of a critical display in which one word serves as the target, one as a competitor, one as a prime and two as distractors. The instructions carrying the prime and target words are illustrated on the left.

On each trial, a display consisting of five labeled buttons was accompanied by two auditory instructions. On critical trials, the first auditory instruction was the carrier sentence for the prime word and the second instruction was the carrier sentence for the target word (presented immediately after the listener selected the prime word on the display). On these trials, the auditory instruction never referred to the competitor or the two unrelated words that were present in the visual display. Also, the minimal pairs for the eight potentially ambiguous prime words (e.g., *clam* for the prime word *clan*) were never present on the visual display nor referred to in the auditory instructions. The procedures for creating the four experimental lists and assigning the four conditions to the words within each list were similar to Experiments 1 and 2. Likewise, the procedure for controlling for the secondary factors of lexical frequency, semantic and phonological similarity and orthographic complexity of the words and lexical ambiguity status of the prime words when assimilated were the same as those applied to Experiments 1 and 2 (refer to section 3.1).

The same set of 72 filler trials and practice trials used in the two identification experiments were used in the current eye tracking experiments. Similar to the critical trials, words in the filler trials were played within their original carrier sentences.

## 4.2 Experiment 3: Nasal Place Assimilation

### 4.2.1 Method

#### 4.2.1.1 Participants

Thirty-two native English speakers were recruited at the University of Toronto. None of the participants in Experiment 3 took part in any of the previous experiments. There were 25 females and seven male adults between 18-32 years of age ( $M = 21$  years). All participants reported no history of hearing or speech difficulty and had either normal or corrected vision. Participants were requested to complete a language background questionnaire and provided a written consent form prior to the experiment. Participants reported native proficiency in English ( $M = 4.94$ ,  $SD = 0.25$ , with 5 being the highest proficiency level), that they had learned English in their early childhood ( $M = 2$  years), and were predominantly using it in everyday life. Four participants were excluded from the final data analysis due to technical problems. All participants received monetary compensation.

#### 4.2.1.2 Materials

The visual display and auditory stimuli were prepared as was described in section 4.1. The critical stimuli used in Experiment 3 specifically consisted of 24 prime and 24 target words that ended in the coronal nasal /n/, either in the canonical coronal form (unassimilated) or assimilated in place of articulation to the following labial consonant. Also, the full set of stimuli included two practice exercises and 72 filler trials. As described in the previous sections, target competitors on the visual display were always minimal pairs of the target words that ended in a labial nasal consonant /m/.

#### 4.2.1.3 Procedure

Participants were tested one at a time. Similar to Experiments 1 and 2, each participant was seated in a sound attenuated booth at approximately 109 cm from a computer screen and two loudspeakers. Before the start of the experiment, the experimenter provided information about the eye tracking device and described the experimental task. Participants were informed that before every new trial a black dot would appear at the centre of the screen for the purpose of calibration and that they need to fixate on it in order to proceed to the next trial. The experimenter then fitted the participant with the eye tracker (sampling rate 250 Hz; EyeLink II,

SR Research, Ottawa, Canada), and conducted a calibration procedure. The experiment always began with two practice trials, followed by the main experiment. At the beginning of each trial, the visual display containing five labeled buttons appeared for 3000 ms to provide enough time to scan the visual display before an orienting tone beep was played while the mouse cursor simultaneously became visible in the centre of the screen. The first instruction (prime) was played immediately afterward. Mouse clicks were only effective after the onset of the word *button*. Immediately after the participant's first click response, the second auditory instruction that carried the target word played automatically and again the participant was required to click on the appropriate button. After the second mouse click, the screen went blank. Five hundred milliseconds after each trial, an eye-drift correction procedure was automatically conducted before proceeding to the next trial. During this process, a black dot would appear on a white display background. Participants were instructed to fixate the dot so that the tracker could recalibrate in case there are any shifts in calculation of the eye movements. This also brings the eye fixation to the centre of the display so the eyes would always be in the same position at the beginning of every trial. Click responses and eye movements during every trial were automatically recorded. Overall, each experimental session took 40 minutes on average from beginning to end.

#### 4.2.2 Results

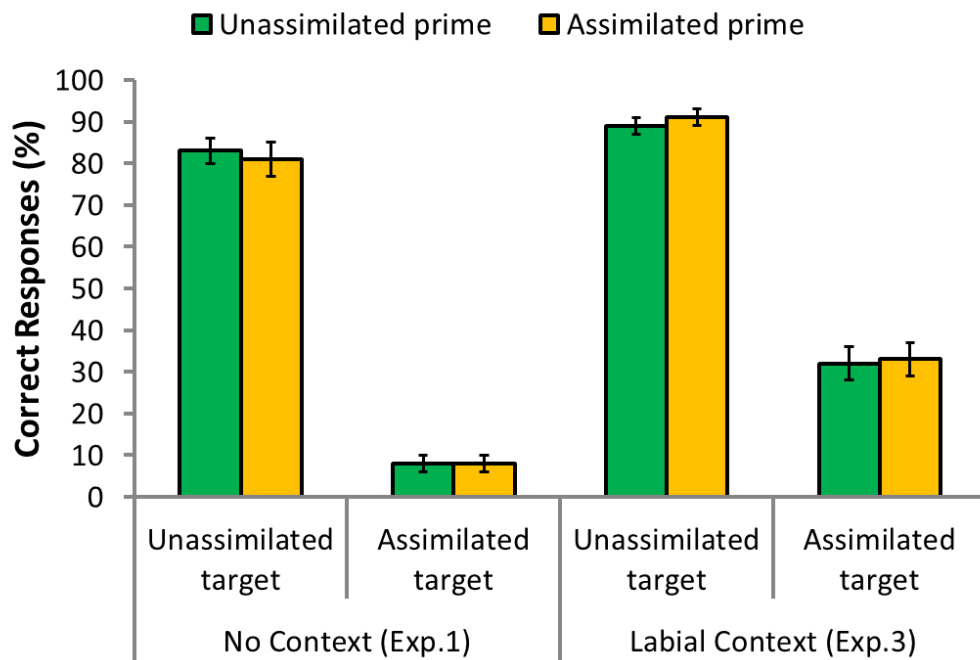
##### *Behavioural data: item selections and mouse click response times*

Participants' mouse click selections for prime and target sentences, as well as the average response times (RT) for the prime sentences, were analyzed in the same way as Experiments 1 and 2.<sup>12</sup> The identification scores in each condition were calculated based on the percentage of the final selections where the prime or target words ending in the coronal nasal were clicked on. For prime sentences, the participants selected the intended word in every trial regardless of the experimental condition. This pattern of results was expected as there were no phonologically and/or orthographically similar words present on the display for the prime items. For target sentences, however, on average the participants correctly identified the intended target word 89%

---

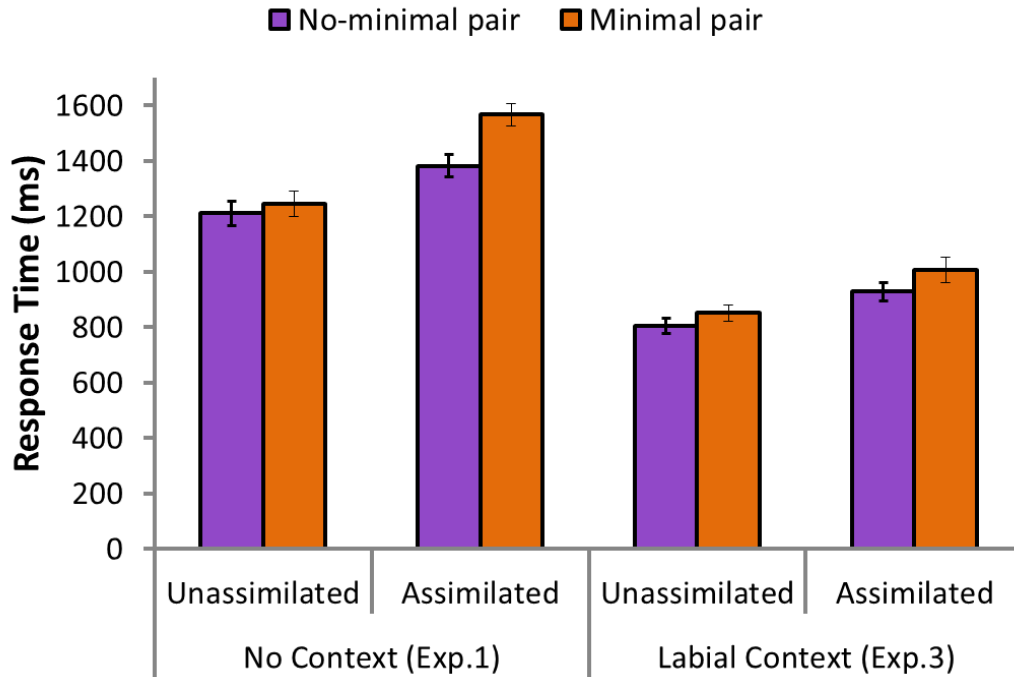
<sup>12</sup> Again, RTs for target sentences were not analyzed due to the variable distance for the mouse to travel from its resting position after clicking on the prime button (see footnote 11).

of the time when the target word was unassimilated and only 32% of the times when it was assimilated (Fig. 4-2, right panel). Compared to Experiment 1 (Fig. 4-2, left panel), where listeners could only hear the isolated words, there was an increase in the average percentage of coronal click responses in Experiment 3, where the phonological context was provided, especially when the target words were assimilated (25% improvement). A generalized linear mixed effects model (*glmer* function from *lmerTest* package in R; Kuznetsova et al., 2016) was conducted on the click response data from Experiments 1 and 3. Fixed effects of Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated), Experiment (Experiment 1 vs. Experiment 3) and their interaction were included. Random intercepts for participant and item and random slopes for Prime and Target by participant were also included. The results showed a significantly higher number of coronal responses when the target words were unassimilated compared to when they were assimilated ( $\beta = 4.45$ ,  $SE = 0.45$ ,  $z = 10$ ,  $p < 0.001$ ). The two experiments were also found to be significantly different, whereby the percentage of coronal responses were overall higher in Experiment 3 compared to Experiment 1 ( $\beta = 1.89$ ,  $SE = 0.41$ ,  $z = 4.57$ ,  $p < 0.001$ ). The main effect of Prime and the interactions were not significant ( $ps > 0.07$ ).



**Figure 4-2.** Percentage of correct responses for target words ending in nasal stops in isolation (left panel, Exp. 1) and in their carrier sentence (right panel, Exp. 3), across conditions. The error bars represent the standard error of the mean.

Additionally, participants' average RT for correct responses to prime sentences was calculated from the offset of the prime word/onset of the word 'button' in the prime carrier sentence "click on the \_\_\_ button.", up to the time point when one of the buttons on the display was clicked on by the participant. Similar to previous experiments, outlier data points were identified and replaced by the average RT ( $\pm 2$  SD from the mean; 3.6% of total observations). RTs for prime words with a minimal pair competitor (e.g., *bean* where *beam* is also a word) and those with no competitor (e.g., *lean*) were calculated for unassimilated and assimilated primes (Fig. 4-3, right panel). A linear mixed effects model using *lmer* function from *lmerTest* package in R was conducted on the average RTs with the fixed effects Assimilation (unassimilated vs. assimilated), Minimal Pair (minimal pair vs. no-minimal pair), Experiment (Experiment 1 vs. Experiment 3) and their interaction. Random intercepts for participant and item were included, as well as random slopes for Assimilation and Minimal Pair by participant. The results of the analysis showed significant main effects of Assimilation ( $\beta = -322.15$ ,  $SE = 44.2$ ,  $t = -7.29$ ,  $p < 0.001$ ) and Minimal Pair ( $\beta = -184.08$ ,  $SE = 51.01$ ,  $t = -3.61$ ,  $p < 0.001$ ). There was also a significant main effect of Experiment indicating that the RTs were significantly slower in Experiment 1 ( $M = 1351$  ms) compared to Experiment 3 ( $M = 897$  ms,  $\beta = -560.3$ ,  $SE = 53.99$ ,  $t = -10.38$ ,  $p < 0.001$ ). There were also significant effects of interaction between Assimilation and Minimal Pair ( $\beta = 150.72$ ,  $SE = 53.01$ ,  $t = 2.84$ ,  $p = 0.005$ ), Assimilation and Experiment ( $\beta = 166.57$ ,  $SE = 58.47$ ,  $t = 2.85$ ,  $p = 0.005$ ) and Minimal Pair and Experiment ( $\beta = 106.08$ ,  $SE = 49.88$ ,  $t = 2.13$ ,  $p = 0.03$ ). The interaction between the three factors was not significant ( $p = 0.09$ ). A series of post-hoc pairwise comparisons of the least square means (*lsmeans* package in R; Lenth, 2016) were conducted. The results showed that in Experiment 3, the listeners' prime RTs were significantly slower when the prime was assimilated ( $M = 967$  ms) compared to when it was unassimilated ( $M = 827$  ms,  $SE = 24.22$ ,  $p < 0.001$ ). Also, RTs in Experiment 3 were faster compared to RTs in Experiment 1, both in unassimilated condition ( $SE = 43.13$ ,  $p < 0.001$ ) and in assimilated condition ( $SE = 45.51$ ,  $p < 0.001$ ). Unlike in Experiment 1, listeners' RTs in Experiment 3 were no longer significantly different between no-minimal pair and minimal pair conditions, regardless of the assimilation condition ( $M = 866$  vs.  $928$  ms,  $p = 0.5$ ).



**Figure 4-3.** Average correct response RT for prime words in isolation (left panel, Exp. 1) or in their carrier sentence (right panel, Exp. 3) that ended in nasals. Plotted based on assimilation condition for primes with or without a minimal pair.

The results of the identification task in the current experiment suggests that the participants had an overall higher preference for selecting a word ending in a labial nasal upon hearing an assimilated word when the following phonological context (i.e., the initial sound of the word *button*) provided a viable context for assimilation. The clear bias to select the words ending in the labial over the coronal nasal in this experiment matches with the patterns observed in the acoustic and perceptual analyses of the current set of stimuli. As shown in Section 2.2, the acoustic analysis of the current stimuli suggested a strong/complete degree of assimilation for the nasal consonants. Additionally, as shown in Experiment 1, the same words were perceived more often as ending in a labial nasal when heard in isolation (refer to Section 3.2). Consequently, in the current assimilated forms, the acoustic cues to labial place of articulation are more prominently perceived by listeners than the cues to coronal place. Crucially however, compared to the results of Experiment 1, the current results demonstrate a 24% improvement in the percentage of (correct) identification of the underlying place of articulation in assimilated word-



final nasals. This indicates the phonological context in fact had a positive influence on the perception of place of articulation in assimilated word-final nasal consonants even though the acoustic cues of the nasals were not in favor of the coronal place of articulation. Based on these patterns, listeners' judgment of the place of articulation of a strongly assimilated nasal stop is clearly modulated by phonological context information, although acoustic cues intrinsic to the sound seem to have relatively more influence on the final identification. The current results, however, did not show any effect of priming for processing assimilation.

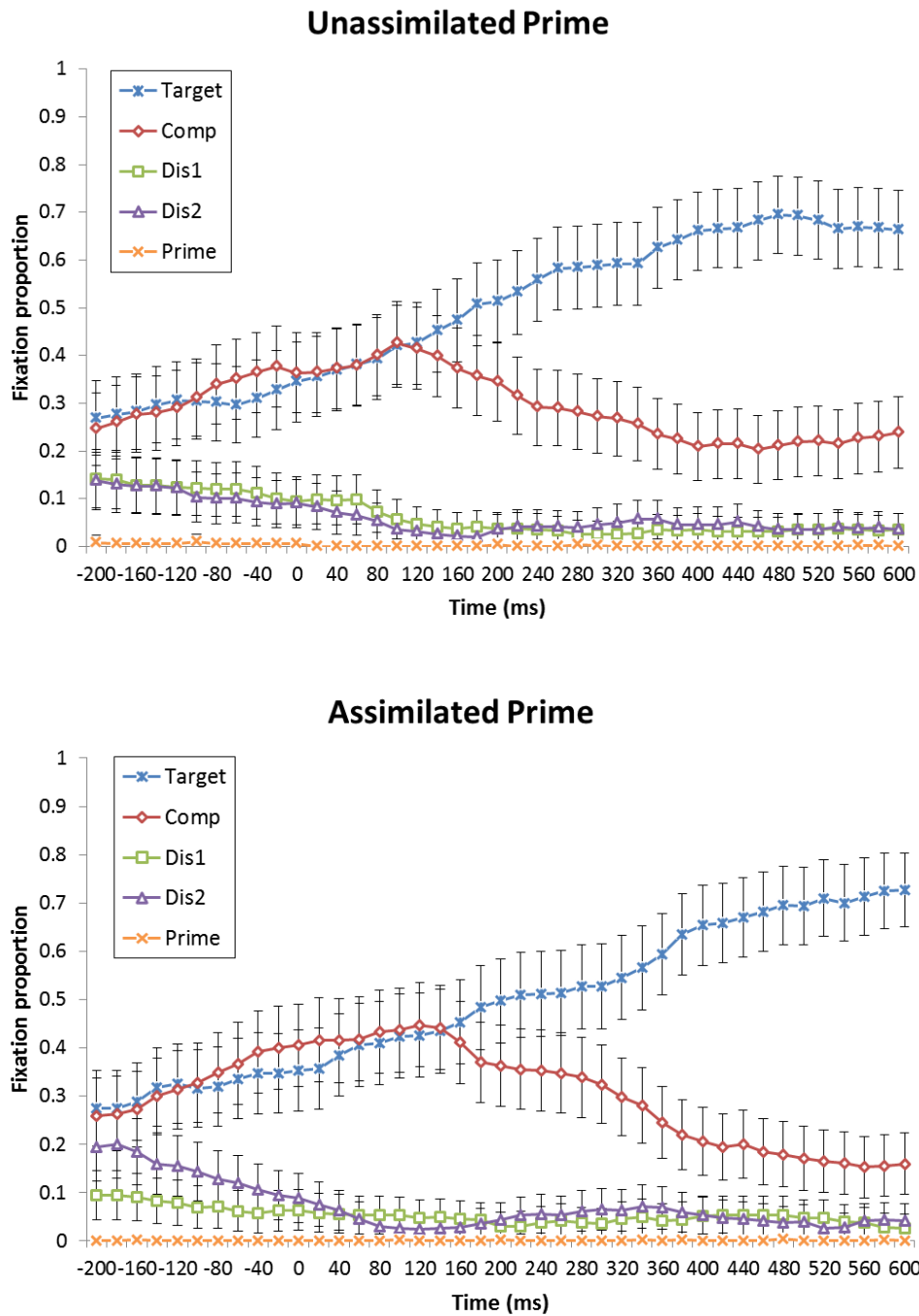
The results also show faster RTs for prime items when the phonological context was provided compared to when the prime words were presented in isolation suggesting that listeners used the contextual information to facilitate accessing lexical items. Also, when the following labial context was made available, the processing time in the assimilated condition was slower compared to the unassimilated condition for both prime words with or without a minimal pair competitor suggesting higher processing load in the former case. Unlike in Experiment 1, where the assimilated primes with a minimal pair competitor were identified slower than primes with no minimal pair, in Experiment 3, the RTs were not affected by this factor. This shows that the availability of the triggering phonological context assists in the process of lexical access by reducing the level of competition that might exist between the perceptually confusing lexical items.

### ***Eye movement data***

Eye-fixations were analyzed in all experimental conditions for effects of place assimilation and priming on processing the target words. The fixation data were aligned at the onset of the consonant [b] at the beginning of the word *button*, which is the phonological context for place assimilation. The data were then extracted from a window extending from 200 ms before the onset of [b] (0 ms on the graphs) to 600 ms after this time point. This window covers the maximum duration of the word-final nasal consonants (160 ms) and the maximum period from the onset of the word *button* until the end of the carrier sentences across all stimuli (516 ms). Any effect associated with the formant transition from the vowel preceding the nasal consonant or any later phonological context effect on processing assimilation should therefore be possible to observe within this time window. The proportion of the eye-fixations to the target, target competitor, prime and the two distractor items was calculated in 20 ms bins within this interval

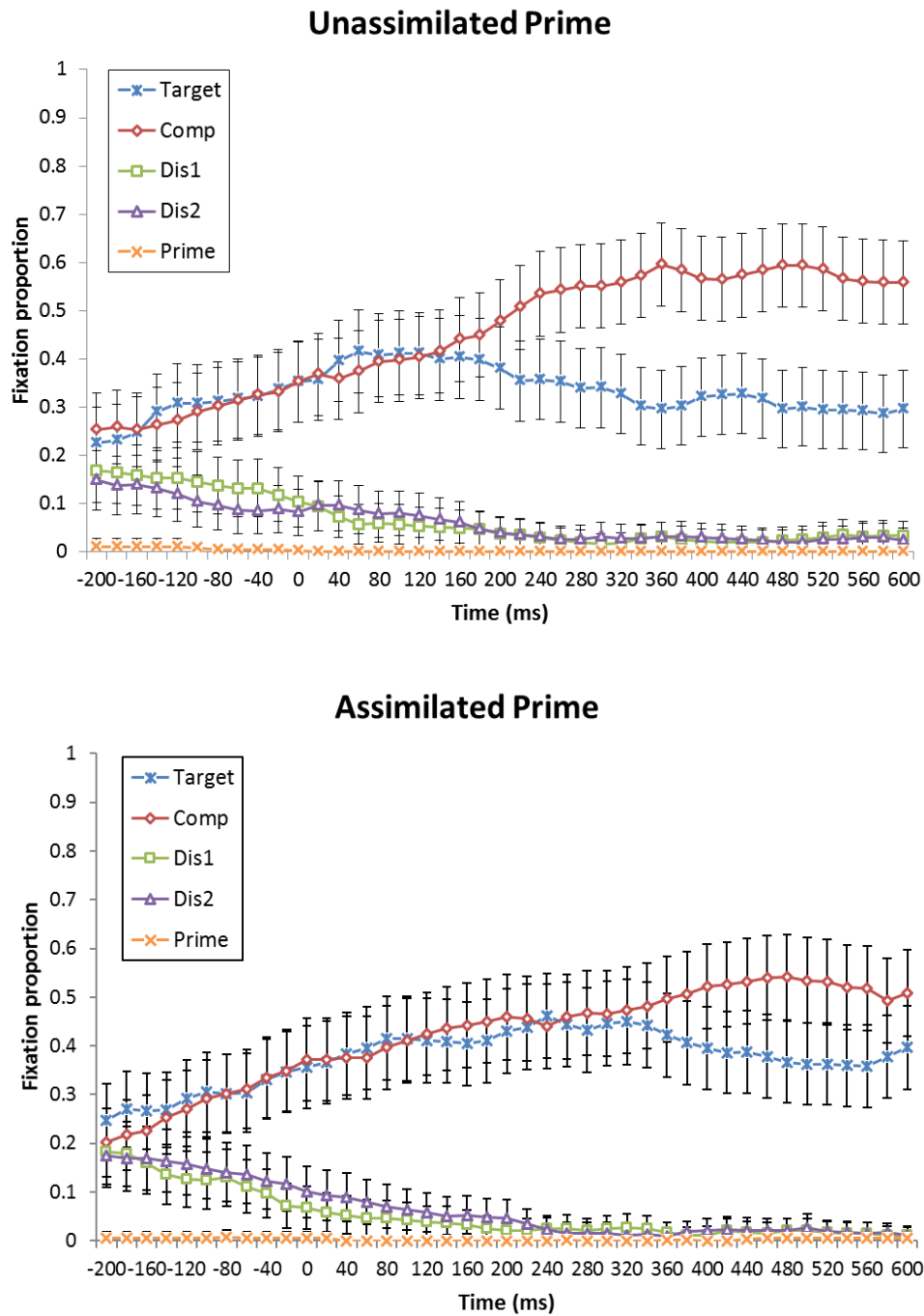
for ease of analysis. The data from one trial (assimilated prime *pin*, unassimilated target *rune*) from one of the participants was missing and, as a result, the trial was removed from the analysis.

The two graphs in Figure 4-4 show the proportion of the participants' fixations to the target (e.g., *dine*) and the competitor (e.g., *dime*) as well as the two distractors (e.g., *glove*, *gulf*) and the prime item (e.g., *lean*) upon hearing unassimilated target words within the specified timeframe. Fixation patterns are depicted when the prime was unassimilated (control; Fig. 4-4, top panel) and when it was assimilated (Fig. 4-4, bottom panel). Note that the fixations do not sum to 100% since the fixation proportions to the areas of the visual display other than the five buttons are not depicted. On these graphs, it appears that fixations to the target and competitor items had already diverged from fixations to the prime or the two unrelated items by around 200 ms prior to the onset of the word *button* (-200 ms on the graphs). Fixations to the target item, however, did not diverge from the competitor item until around 200 ms after the onset of *button* (in "Now click on the word \_\_\_ *button*."). Considering the average time for planning and launching an eye movement, which is estimated to be around 150-200 ms (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Magnuson, Dixon, Tanenhaus & Aslin, 2007), the point of divergence approximately corresponds with the earliest time when the acoustic information related to the sound [b] in *button* becomes available. A comparison between the two figures does not indicate a difference in the pattern of fixations based on the presence or absence of assimilation in the priming sentence.



**Figure 4-4.** Mean fixation proportion on buttons associated with the unassimilated targets ending in a nasal, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). The zero mark on the x-axis corresponds to the onset of the following phonological context (/b/ in *button*).

Figure 4-5 depicts the average proportion of fixations to the target, competitor, prime and the two distractor items in the assimilated target condition when the target was preceded by an unassimilated prime (control, top panel) or an assimilated prime (bottom panel). As before, fixations to the unrelated items (i.e., prime and the two distractors) diverged from fixations to the target and competitor items early on, as these words did not share the same onset or vowel with either of the target or competitor words. When the prime was unassimilated (top panel), fixations to the target and competitor items began to diverge from one another approximately around 200 ms, again corresponding to the point in time where the acoustic information associated with the sound [b] in the word *button* becomes available (considering a latency of around 150-200 ms for launching the eye movements). The advantage of the fixations to the competitor over the fixations to the target item in the condition where targets were assimilated closely follows the pattern of the identification results in this condition (68% labial responses). The pattern of fixations when the prime was assimilated (bottom panel), on the other hand, indicates that the competitor advantage observed with the control prime condition was no longer present when the prime sentence contained an assimilated nasal.



**Figure 4-5.** Mean fixation proportion on buttons associated with the assimilated targets ending in a nasal, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). The zero mark on the horizontal axis corresponds to the onset of the following phonological context (/b/ in *button*).

A target advantage score was calculated based on the difference between the probability of fixations to the target and the competitor item (plotted in Figure 4-12 across all participants). The range of the difference is between -1 (always looking at the competitor) and 1 (always looking at the target), and zero means no preference for either word over the time window. For example, for a participant whose probability of fixation to the word *dean* (target) in unassimilated prime and unassimilated target condition was 0.81 and to the word *deem* (competitor) was 0.15, the target advantage would be 0.66. All analyses were performed on the data corresponding to the time when the phonological context became available (at 200 ms, including the delay) up to the end of the time window (at 600 ms). A linear mixed effects model using *lmer* from *lmerTest* in R was performed on the target advantage measure. Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated) and their interaction were included as fixed effects. Random intercepts were included for participant and item, as well as random slopes for Prime and Target and their interaction by participant and random slopes for Prime and Target by item. The results of the tests indicated a significant main effect of Prime ( $\beta = -0.15$ ,  $SE = 0.07$ ,  $t = -2.13$ ,  $p = 0.04$ ) and Target ( $\beta = 0.48$ ,  $SE = 0.09$ ,  $t = 5.12$ ,  $p < 0.001$ ). No significant interaction effect was observed ( $p = 0.1$ ). Although there was no interaction effect, the results of a planned comparison using least square means (*lsmeans* in R) suggested that there were significantly more fixations to assimilated target items that followed an assimilated prime sentence compared to the fixations to assimilated target items that followed an unassimilated prime sentence ( $SE = 0.07$ ,  $t = 2.05$ ,  $p = 0.04$ ). When the target word was unassimilated, however, there was no significant difference in the target advantage between assimilated and unassimilated prime conditions ( $SE = 0.07$ ,  $t = -0.02$ ,  $p = 0.98$ ).

The effect of the assimilated prime on the relative consideration of the target suggests an effect of priming of the assimilation process. This can be observed in the pattern of fixations to the assimilated targets. When the prime sentence was unassimilated (top panel), stronger consideration of the competitor is observed immediately around the offset of the nasal consonant/onset of the word *button* (i.e., 200 ms) despite the upcoming viable phonological context for assimilation, presumably due to the strong match of the acoustic information of the sound with the labial place of articulation. Crucially, when the prime sentence contained assimilation (bottom panel), the competitor was not significantly fixated more than the target throughout the analysis window (i.e., from 200 to 600 ms). This suggests that undoing

assimilation in the preceding utterance influenced recognition of the subsequent assimilated word such that the perceptual mechanism allowed a word containing an underlying coronal nasal to continue as a possible match for the input signal despite the strong acoustic mismatch.

## 4.3 Experiment 4: Stop Place Assimilation

### 4.3.1 Methodology

#### 4.3.1.1 Participants

Thirty-two native English speakers, who did not participate in any of the previous experiments, were recruited from the University of Toronto community and were paid for their participation. 20 participants were female and 12 were male adults between 18-29 years of age ( $M = 20$  years). None of the participants reported having a history of hearing or speech difficulty and had either a normal or corrected vision. Participants completed a language background questionnaire and provided a written consent form prior to the experiment. All participants self-reported learning English in their early childhood ( $M = 0.5$  years) and using it predominantly in their everyday communication. Participants also self-reported native proficiency in English using a scale of 1-5, with 5 being the highest proficiency level ( $M = 4.94$ ,  $SD = 0.25$ ). Four participants were excluded from the final data analysis due to technical problems. All participants received monetary compensation.

#### 4.3.1.2 Materials

The visual display used in Experiment 4 was the same as in Experiment 2. The general format of the auditory stimuli was the same as for Experiment 3, using the same recorded files described in section 4.1. The auditory stimuli consisted of 24 prime and 24 target words in the original carrier sentences (e.g., prime sentence: “Click on the *fit* button.”, target sentence: “Now, click on the *mat* button.”). However, in critical trials the prime and target words all ended in the coronal stop consonant /t/ (either assimilated or unassimilated). The target competitor words (presented as one of the labeled buttons) were always words ending in the labial stop consonant /p/. The two practice and 72 filler words used in Experiment 2 were also included in the corresponding carrier sentences and were used in Experiment 4.

### 4.3.1.3 Procedure

The experimental procedure followed the same format as of Experiment 3. Participants were seated in the sound attenuated booth. After the experimenter's explanations, the eye tracker headpiece was placed on the participant's head, and the calibration procedure was initiated shortly after. Upon finishing the calibration procedure and the first two practice trials, the main experiment began. Similar to Experiment 3, the task was to use the computer mouse to click on the button that best matched with the word mentioned in the auditory instructions. Mouse click responses, RTs for the prime instructions, and eye movement data were collected for each critical trial.

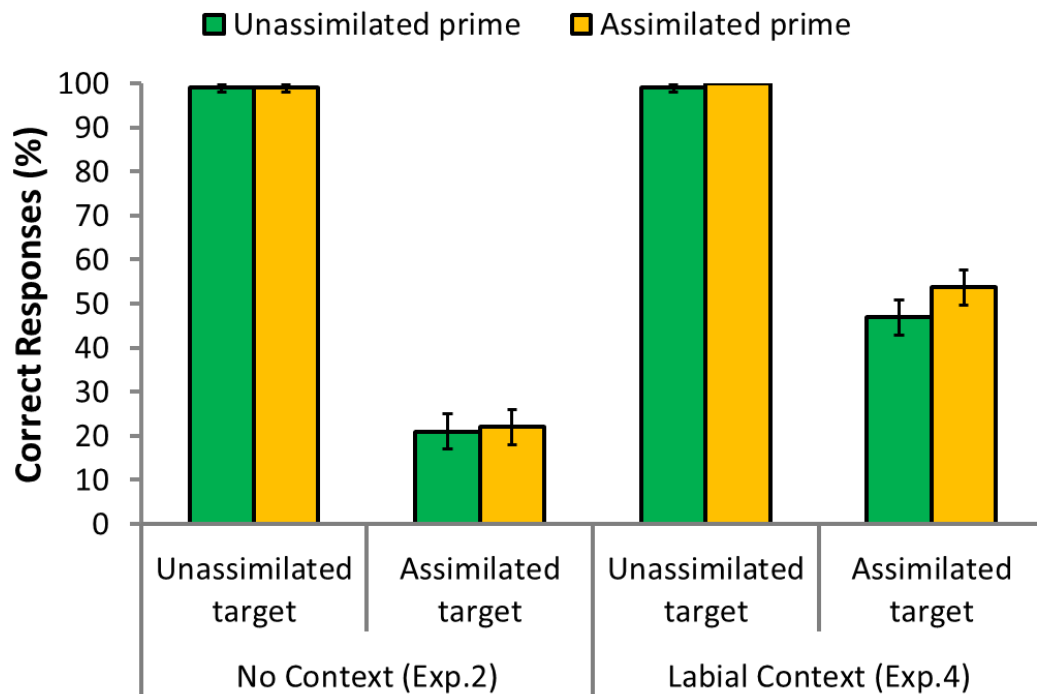
### 4.3.2 Results

#### *Behavioural data: item selections and mouse click response times*

The average percentage of correct identifications for the target words in each experimental condition is shown in Figure 4-6. An examination of the results indicates that the listeners were 100% accurate in identifying the prime words ending in a coronal stop upon hearing the first instruction, regardless of whether the prime was or was not assimilated. As before this is not surprising as there is no similar-sounding word present in the visual array. Around the same percentage of correct selections was obtained for target words in the second instruction when they were unassimilated (99%). When the target words were assimilated, however, listeners' percentage of accurate responses decreased to an average of 50% correct: 47% when the prime was unassimilated and 54% correct when the prime word was also assimilated. When compared to Experiment 2 (no triggering context present), listeners in fact showed an average 29% improvement in percentage of correct click responses. A generalized linear mixed effects model (*glmer* from *lmerTest* in R) was conducted on the click response data from Experiments 2 and 4, with Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated), Experiment (Experiment 2 vs. Experiment 4) and their interaction as fixed effects. Random intercepts by participant and item and random slopes for Prime and Target by participant were also included. The results showed a significantly higher number of coronal responses when the target words were unassimilated compared to when they were assimilated ( $\beta = 10.59$ ,  $SE = 2.72$ ,  $z = 3.9$ ,  $p < 0.001$ ). There was also a significant difference between the two experiment types, where the percentage of coronal responses was overall higher in Experiment 4 compared to Experiment 2



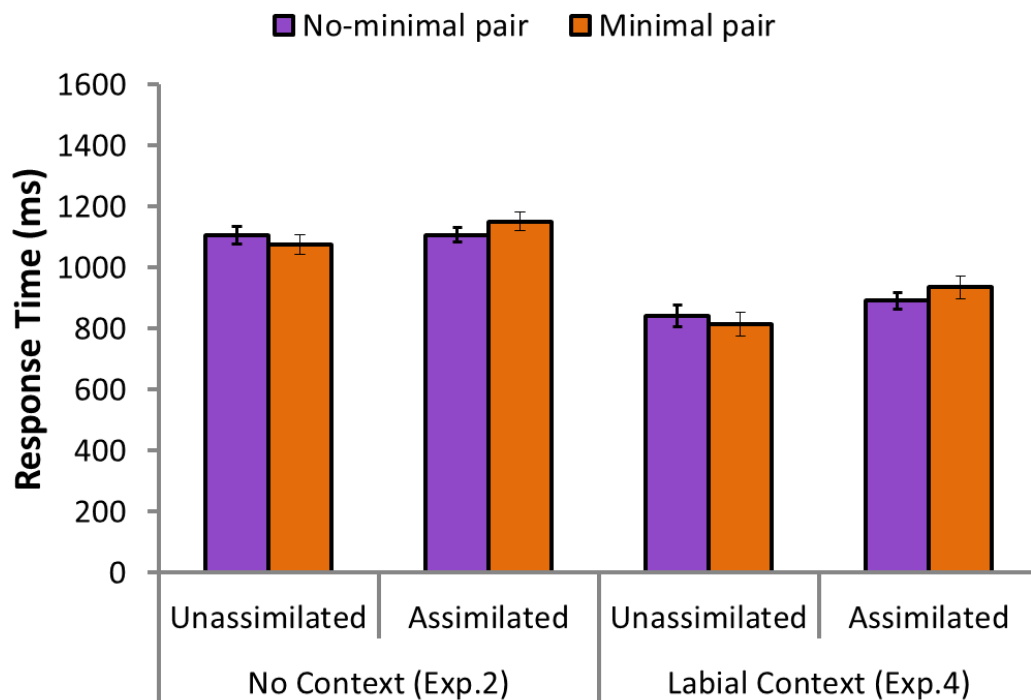
( $\beta = 1.68$ ,  $SE = 0.31$ ,  $z = 5.33$ ,  $p < 0.001$ ). Neither the main effect of Prime nor the interactions were significant (all  $p$ 's  $> 0.56$ ).



**Figure 4-6.** Percentage of correct responses for isolated target words (right panel, Exp. 2) and words in their carrier sentences (left panel, Exp. 4) ending in oral stops, across conditions. The error bars represent the standard error.

Response times for correctly identified prime words were also calculated from the offset of the prime word/onset of the word *button* up to when the corresponding button on the display was clicked on. RTs for the target sentences were not calculated (refer to Section 3.2.2, p. 55). From the total number of data points, 3.9% were identified as outliers ( $\pm 2SD$  from the mean RT) and were replaced by the average RT. Figure 4-7 (right panel) depicts the average RTs in Experiment 4 for unassimilated or assimilated prime words that have no minimal pair competitor (e.g., *fit*) or those that have a minimal pair competitor (e.g., *coat* with the minimal pair *cope*). On the left panels, the corresponding RT results from Experiment 2 are depicted as well. A linear mixed effect analysis (*lmer* from *lmerTest* in R) was conducted on the RT results with Assimilation

(unassimilated vs. assimilated), Minimal Pair (no minimal pair vs. minimal pair) and Experiment (Experiment 2 vs. Experiment 4) and their interaction as fixed effects. Random intercepts for participants and items were included. Only random slopes for Assimilation and Minimal Pair by participant were included. The results showed a significant main effect of Experiment ( $\beta = -215.46$ ,  $SE = 47.81$ ,  $t = -4.51$ ,  $p < 0.001$ ) indicating overall slower RTs in Experiment 2 ( $M = 1110$  ms) compared to Experiment 4 ( $M = 870$  ms). There was a marginally significant main effect of assimilation that suggests a slight increase in RTs when the prime words were assimilated ( $M = 1020$  ms) compared to when they were not assimilated ( $M = 959$  ms) across the two experiments ( $\beta = -73.54$ ,  $SE = 38.29$ ,  $t = -1.92$ ,  $p = 0.06$ ). There was no significant main effect of the minimal pair condition ( $p = 0.2$ ) or any of the interaction terms (all  $p$ 's  $> 0.1$ ).

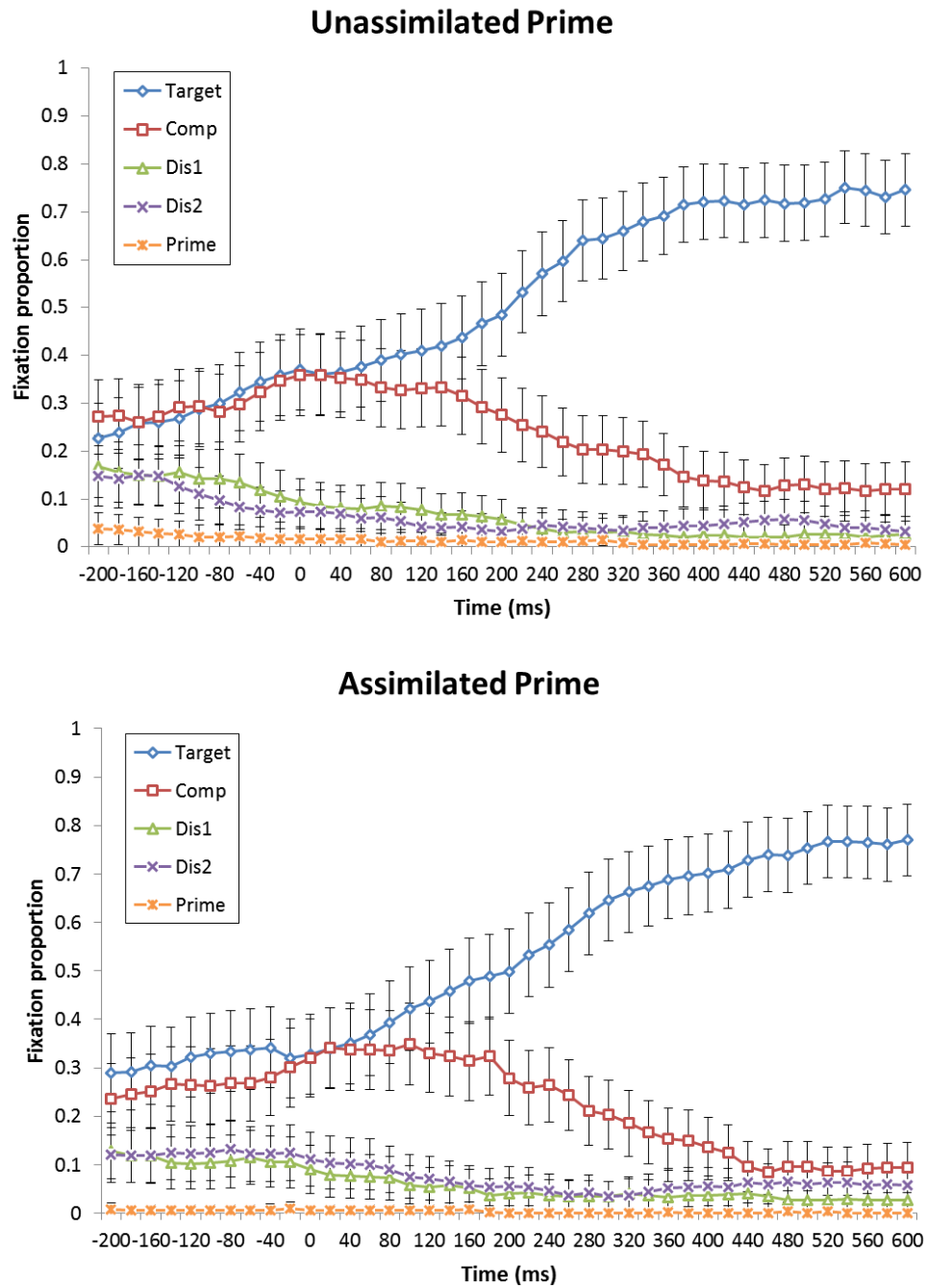


**Figure 4-7.** Average correct response RT for prime words ending in oral stops in isolation (left panel, Exp. 2) or in their carrier sentence (right panel, Exp. 4). Plotted based on assimilation condition for primes with or without a minimal pair.

The current identification results for target words ending in oral stops suggest no effect of priming for processing assimilation. Also, similar to Experiment 2, having a minimal pair competitor did not seem to affect the RT results for prime words ending in an oral stop consonant. Compared to the results of Experiment 2, however, where the triggering context was absent, results of the current experiment suggest an overall improvement in participants' performance in processing assimilated words ending in oral stops. When the triggering context was available, there was a 29% increase in correct identification of the assimilated target words, and the RTs for prime words were on average 257 ms faster compared to the condition when the words were presented in isolation. The overall faster RTs for prime words when the context was made available suggest that contextual information was used to facilitate accessing the lexical items. Importantly, the bias toward choosing the word ending in a labial stop consonant observed in identification results in Experiment 2 turned into an uncertainty about the assimilated consonant place of articulation upon provision of the phonological context. This is a strong indicator for the effect of phonological context on perception of assimilated forms even though the context was insufficient to resolve the ambiguity in favour of the intended target.

### ***Eye movement data***

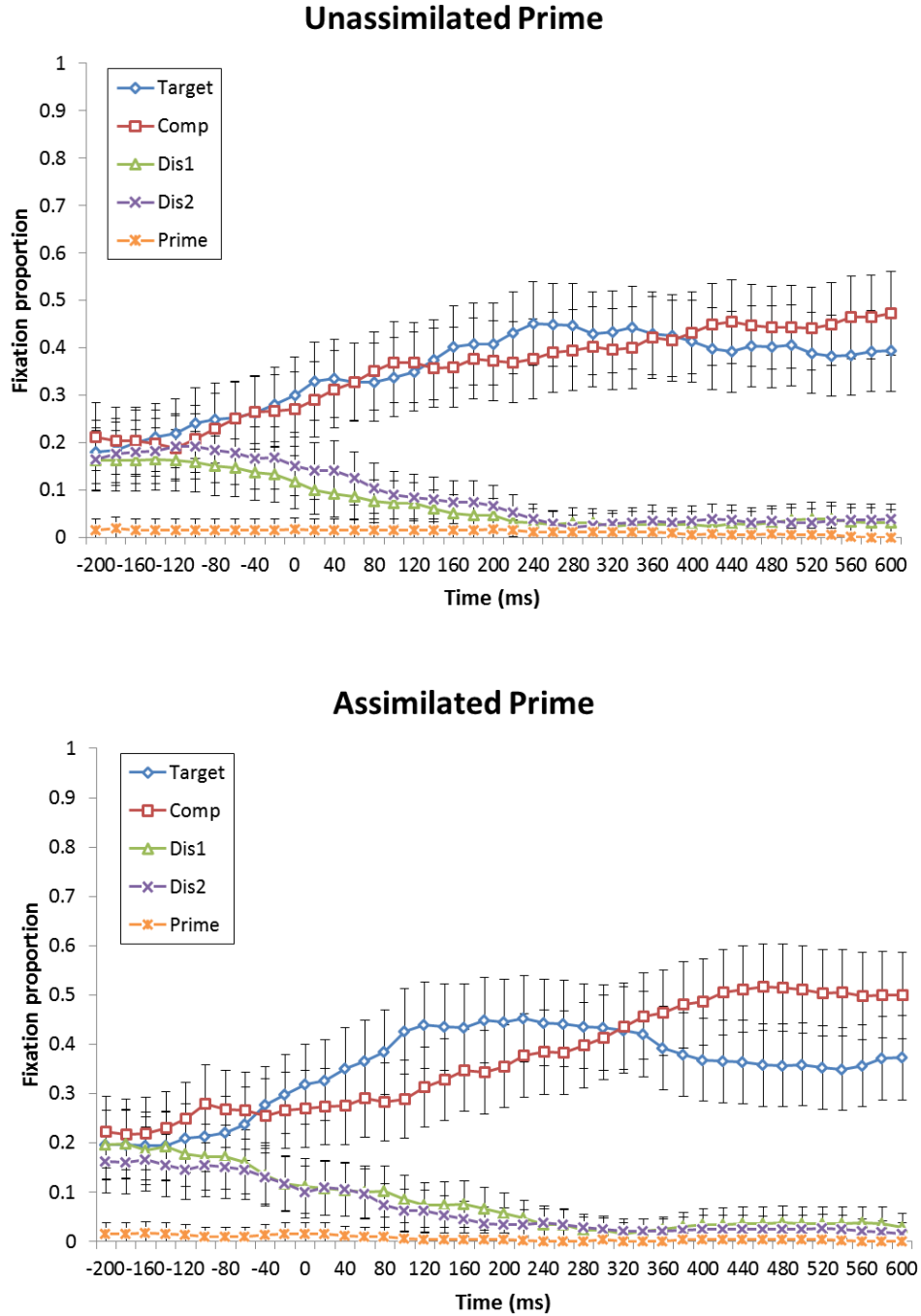
To test for any effect of place assimilation and priming on the word recognition processing, the probability of fixations to the target items ending in a coronal stop (e.g., *cat*) was plotted in relation to the competitor items (e.g., *cap*) as well as the prime (e.g., *root*), distractor 1 (e.g., *dodge*) and distractor 2 items (e.g., *freak*). Average fixation proportions were calculated in 20 ms bins. Similar to Experiment 3, the analysis window extended from 200 ms before to 600 ms after the onset of the consonant [b], located at the beginning of the word *button* (in “Now click on the word \_\_\_ *button*.”), which is where the phonological context for place assimilation becomes available. This time window was chosen to make sure that the period of time associated with any effects of formant transition cues as well as the assimilation-triggering phonological context is included. This is because the maximum duration of the final oral stops across all target and prime words was measured to be 148 ms and the maximum duration from the onset of the word *button* until the end of the carrier sentences across all the conditions was measured to be 504 ms, therefore, falling within the selected time window.



**Figure 4-8.** Mean fixation proportion to buttons associated with the unassimilated targets ending in an oral stop, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). The zero mark on the horizontal axis corresponds to the onset of the following phonological context (/b/ in *button*).

Fixation patterns for the condition with unassimilated target words are depicted in Figure 4-8. As shown in the plots, fixation rates to the target and competitor items diverged from fixations to the prime or the two unrelated items well before the onset of the word *button*, which corresponds to point 0 on the x-axis. Stronger consideration of the target item over the competitor item in either of the prime assimilation conditions started around 160 ms after the onset of [b] in the word *button*. Since it takes around 150-200 ms to plan and launch an eye movement, this means that the target advantage started to take place around the time when the acoustic information related to the sound [b] in *button* became available.

Fixation patterns in the assimilated target condition are depicted in Figure 4-9 for conditions with an unassimilated prime (control, top panel) or an assimilated prime (bottom panel), respectively. Similar to the findings in the unassimilated target condition, fixations to the target and competitor items already diverge from fixations to unrelated items (i.e., prime and the two distractors) prior to the onset of [b]. On the other hand, no consistent preference for the target over the competitor or vice versa can be observed up to the end of the trial epoch. This pattern is generally predicted due to the fact that the participants' final selections were around chance level, indicating perceptual confusion. Despite this general pattern, however, there might be a slight advantage of the target fixations over the competitor item around 40 ms prior to the onset of *button* (-40 ms) to 320 ms after, when the prime was assimilated (Fig. 4-9, bottom panel). The target advantage approximately corresponds with the time when the acoustic information related to the word-final stop consonant became available, suggesting a possible trend for an earlier phonological priming effect. However, this pattern was reversed after this time window, during the time when the acoustic information associated with [b] in *button* became available.



**Figure 4-9.** Mean fixation proportion on buttons associated with the assimilated targets ending in oral stops, target competitor, prime and two distractors over time when the prime word was unassimilated (upper panel) or assimilated (lower panel). The zero mark on the horizontal axis corresponds to the onset of the following phonological context (/b/ in *button*).

A mixed effects model, using *lmer* from *lmerTest* in R, was performed on the target advantage measure from the time when the phonological context became available (200 ms including the delay) up to the end of the time window (600 ms). The fixed effects were the Prime (unassimilated vs. assimilated) and Target (unassimilated vs. assimilated) and their interaction. Random intercepts for participant and item and random slopes for Prime and Target and their interaction by participant were included. The results of the tests indicated a significant main effect of Target ( $\beta = 0.59$ ,  $SE = 0.07$ ,  $t = 9.01$ ,  $p < 0.001$ ). The main effect of the Prime was not significant ( $p = 0.3$ ), nor was the effect of the interaction ( $p = 0.4$ ). To test the trend that was observed in the advantage of the target fixations over the competitor item when the prime and target were both assimilated, the same mixed effects model was used on the target advantage measure across all conditions, this time for a time window from -40 to 320 ms. Again, the results showed a significant main effect of Target ( $\beta = 0.15$ ,  $SE = 0.07$ ,  $t = 2.27$ ,  $p = 0.03$ ) and no effect of Prime ( $p = 0.7$ ) or an interaction effect ( $p = 0.8$ ), indicating that the trend for the priming effect was in fact nonsignificant.

#### 4.4 Cross-Experiment Comparisons and General Discussion

Experiments 3 and 4 investigated the recognition of assimilated words ending in coronal nasal or oral stop consonants when the following phonological context that would trigger assimilation was present. Of additional interest was the effect of listeners' recent experience of processing an assimilated word to examine the possibility of observing any differences in processing of assimilation in nasals versus oral stops. For this purpose, a method was used that combined eye tracking with a process priming paradigm. To achieve the goals, identification scores, prime response times and patterns of eye movements were analyzed.

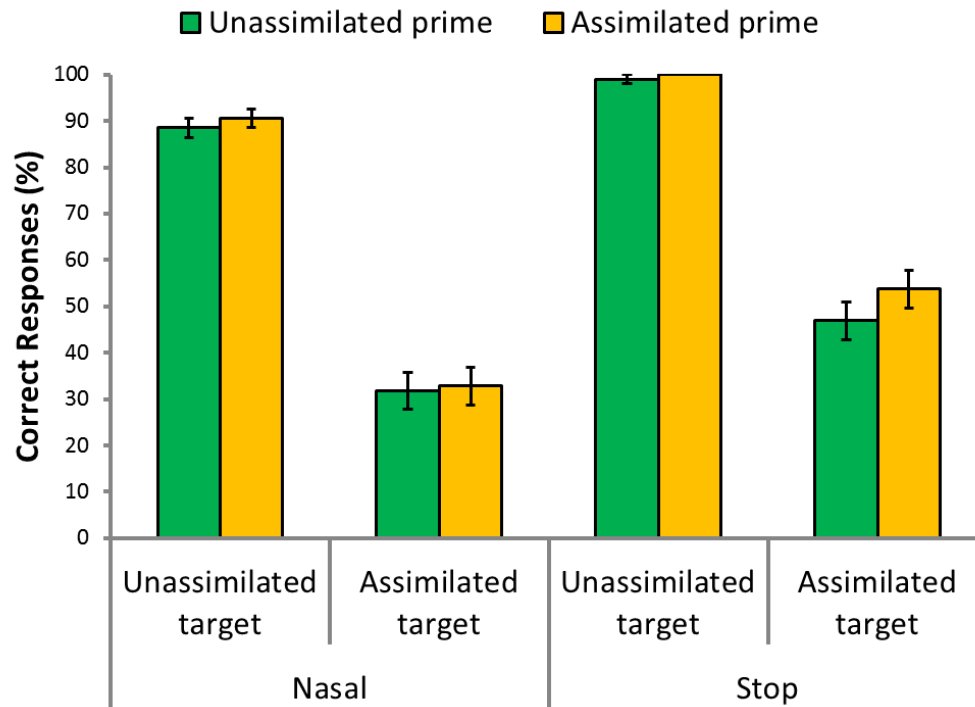
The results from Experiment 3 showed that words ending in strongly/completely assimilated nasal consonants (e.g., *dine<sup>m</sup>*) were most often confused with the words ending in a labial nasal instead (e.g., *dime*), even though the following phonological context was present and viable for place assimilation to occur. A comparison between the results of Experiments 1 and 3 (i.e., words in isolation versus words in the original phonological context), however, revealed that in fact the presence of a viable phonological context resulted in an increase (~ 24% increase) in correct identification of the target words when they were assimilated. The same effect of the phonological context on recognition of assimilated forms was observed for words ending in

assimilated stop consonants (i.e., Experiment 4, compared to Experiment 2). Here again, provision of a viable phonological context for assimilation resulted in an increase in identification rate of the word-final assimilated stops as coronal rather than labial (~29% increase). Note to the fact that the correct identification rate of words ending in assimilated oral stop consonants at around 50% does not simply reflect chance level, rather it indicates that listeners were compensating after the context became available compared to when it was not available (from 21% to 50% correct). These results suggest a relatively constant improvement in identification rates across consonant categories (i.e., nasal and oral stop consonants). In addition, the results indicate an improvement even in case of complete/strong place assimilation, particularly for words ending in nasal consonants where the acoustic cues to place of articulation favor a labial interpretation (Fig. 4-10). This provides strong direct evidence for the positive effect of the phonological context on the recognition of assimilated words and further reveals that such an effect on listeners is constant regardless of the degree of assimilation and the category of assimilated consonants. These results are unlike what has been suggested in previous studies focusing on the viability effect of the context, where the effect of a viable phonological context (as opposed to an unviable context) disappeared in cases where the assimilation was strong/complete. These findings are overall in accordance with certain theoretical accounts of compensation for assimilation such as perceptual integration, feature parsing and phonological inference, which consider a role for the triggering phonological context in recognition of assimilated words. An underspecification account, on the other hand, would predict an identification rate at or better than chance level for words either ending in assimilated nasal or oral stops even when the context is not provided. This is because, according to this account, such assimilated words do not mismatch either a labial or an underspecified coronal place of articulation and therefore both lexical forms should be activated, regardless of whether the assimilation-triggering context is available or not. Moreover, the relatively constant effect of context regardless of consonant type or degree of assimilation also indicates that compensation for assimilation does not solely rely on acoustic information from the assimilated consonant and the following triggering sound, as was proposed by accounts such as feature parsing and perceptual integration. These accounts would predict variation in the effect of context depending on the acoustic characteristics of the assimilated consonant, yet the effect in the current experiments was very similar for both oral and nasal stops



In the unassimilated condition, the presence of the following phonological context also improved identification scores for word-final nasals, suggesting that lower-level acoustic information from the following sound helped with the perception of place of articulation in this sound class.

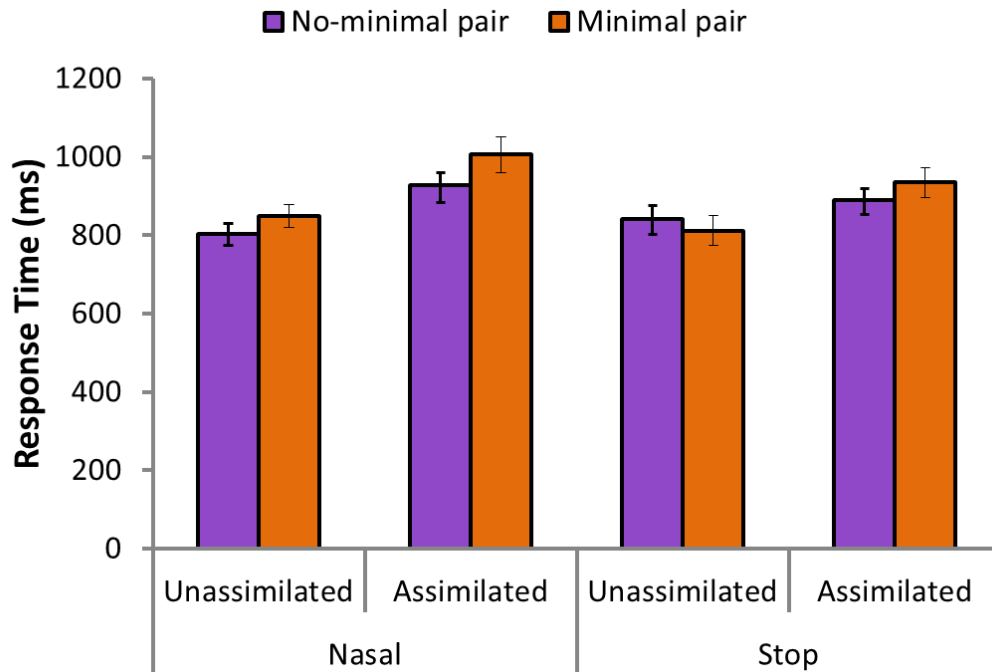
A logistic mixed effects model (using *glmer* from *lmerTest* in R) was conducted on the click response data from Experiments 3 and 4 (words ending in nasals and words ending in stops respectively). Fixed effects were Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated), Consonant Type (nasal vs. stop) and the interaction between Prime and Consonant Type and Target and Consonant Type. Random intercepts for participant and item as well as random slopes for Prime and Target by participant were included. Results showed that, as expected, listeners selected words whose final sound had a coronal place of articulation significantly more often when they heard unassimilated target words compared to when they heard assimilated targets ( $\beta = 3.41$ ,  $SE = 0.28$ ,  $z = 12.31$ ,  $p < 0.001$ ). There was also a significant difference between the percentage of coronal mouse click responses when compared between the words ending in nasal versus stop consonants, where listeners overall selected coronal place of articulation more often upon hearing the words ending in stop consonants ( $\beta = 1.04$ ,  $SE = 0.36$ ,  $z = 2.88$ ,  $p = 0.004$ ). There was also an interaction effect between the assimilation condition of the target words and type of consonants ( $\beta = 2.27$ ,  $SE = 0.77$ ,  $z = 2.95$ ,  $p = 0.003$ ). Post-hoc analysis using *lsmeans* in R on the interaction term revealed that, regardless of prime assimilation condition, the percentage of coronal selection for words ending in oral stops was higher than for words ending in nasals when the target words were unassimilated (99% vs. 90%;  $SE = -3.19$ ,  $p < 0.001$ ) or assimilated (50% vs. 32%;  $SE = -0.92$ ,  $p = 0.02$ ). The overall lower number of correct responses particularly in unassimilated condition for words ending in nasals compared to stop consonants is consistent with the claim that acoustic cues to place of articulation are perceptually less salient for nasal than for stop consonants (e.g., Ohala & Ohala, 1993).



**Figure 4-10.** Percentage of correct responses for target words ending in nasals (left panel, Exp.3) or oral stops (right panel, Exp. 4) when presented in their carrier sentence. Plotted based on the assimilation condition for prime and target words. The error bars represent the standard error.

The results of the analysis on response times (RTs) for prime words ending in nasal and stop consonants in Experiments 3 and 4 are depicted in Figure 4-11. To compare results between the two experiments, a linear mixed effects model (*lmer* from *lmerTest* in R) was conducted on the RT data from the two groups. Fixed effects were Prime (unassimilated vs. assimilated), Target (unassimilated vs. assimilated), Consonant Type (nasal vs. stop) and their interaction. Random intercepts by participant and item were included in the model. The results indicated a significant main effect of Assimilation: Listeners RTs were overall slower when the prime words were assimilated ( $M = 940$  ms) compared to when the primes were unassimilated ( $M = 827$  ms) regardless of the final consonant ( $\beta = -155.58$ ,  $SE = 35.22$ ,  $t = -4.41$ ,  $p < 0.001$ ). This indicates that assimilation results in an extra processing load for the recognition system. There were however no significant main effects of Minimal Pair ( $p = 0.09$ ) or Consonant Type ( $p = 0.3$ ). Also, none of the interactions were significant (all  $p$ 's  $> 0.4$ ). These results suggest that the

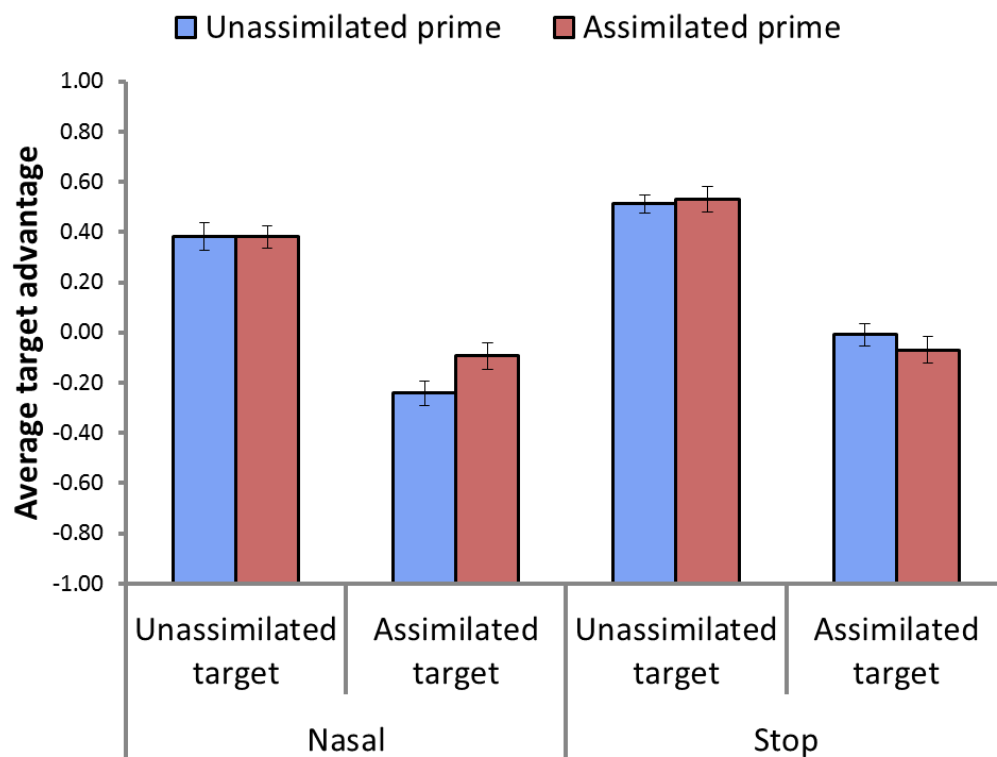
difference observed between RTs for isolated assimilated prime words ending in nasals versus oral stops (Section 3.4), where assimilated prime words ending in nasals were identified slower than those ending in oral stops (especially if they had a minimal pair competitor), is no longer present when the triggering context for assimilation is provided for the listeners. This further indicates that, in addition to the overall facilitatory effect of the context on recognition of assimilated forms, prime words ending in nasal stops potentially benefit from the presence of the context in terms of reducing processing load and lexical competition even more than prime words ending in oral stops.



**Figure 4-11.** Average correct response RT for prime words that end in nasals (left panel) or oral stops (right panel) when presented in their carrier sentence. Plotted based on assimilation condition for primes with or without a minimal pair.

Finally, an examination of the eye movement data revealed a process priming effect for the recognition of assimilated words ending coronal nasal consonants. Specifically, when listeners had just correctly recognized an assimilated word in the previous (prime) sentence, they fixated significantly more on the (coronal) target word as soon as the following phonological context became available compared to a control condition in which the prime word was unassimilated

(Fig. 4-12). This priming effect, however, was not observed for assimilated stop consonants. An examination of the target advantage in the conditions with *unassimilated* targets also did not show a priming effect for either nasal or stop consonants. The advantage of the target fixations over the competitor items for both nasal and oral stops in this condition started approximately at the time when the following phonological context information became available (around 160-200 ms after the onset of *button*, refer to Fig. 4-4 and Fig. 4-8).



**Figure 4-12.** Average advantage of fixations to the target over the competitor words for words ending in nasals (left panel, Exp. 3) or ending in oral stops (right panel, Exp. 4), across conditions (within 200-600 ms after the onset of /b/ in *button*).

To summarize these cross-experiment analyses, the mouse click data provides robust evidence in favor of the effect of the phonological context on perception of assimilated forms for both nasal and oral stop consonants. This effect was not strong enough, however, to result in a complete compensation for assimilation in either of the sound groups, perhaps due to the strong degree of

assimilation in the current stimuli. This is despite the fact that, in the current set of experiments, the following phonological context was always predictable (the word *button*), therefore, minimizing the possibility of misperception of the context. These results are in accordance with the results of the previous studies conducted by Gaskell and colleagues (Gaskell & Marslen-Wilson, 2001; Gaskell & Snoeren, 2008), who found a lack of complete compensation for strongly/completely assimilated forms suggesting higher level information (e.g., semantic context) is required for fully recovering the underlying forms. In addition, measures of online processing demonstrate a divergence between the two sound categories regarding the perceptual mechanisms involved. A “process priming” effect (as opposed to a form priming effect, which would be most apparent when the previous word was *unassimilated*) was observed for assimilated nasal consonants but not for assimilated oral stops. This effect occurred at the point where the triggering phonological context was encountered. This indicates that the contextual information was processed differently for the two sound types.

Considering the relatively higher degree of misperception of place of articulation for the word-final nasal consonants and the dissociation that was found between recognition of words ending in unassimilated or assimilated nasal versus oral stops, one intriguing question is whether such distinctions are also reflected in the structure of the English lexicon. Are potentially ambiguous minimal pairs that end in a labial or coronal oral stop or nasal consonant (the group of lexical items that were the focus of this study) equally frequent in the lexicon? How similar are such perceptually confusing words to each other in terms of lexical characteristics such as semantic relatedness, which can influence the likelihood of words occurring in the same context? The next chapter investigates the answer to these questions as a means to provide a better understanding of language processing and its relationship with linguistic patterns.

## Chapter 5

### Corpus Study: Minimal Pairs Ending in /n/-/m/ or /t/-/p/

Models of spoken word recognition generally agree that the process of accessing a lexical item involves competition of multiple lexical candidates that best match with the bottom-up acoustic input (Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 2001). Furthermore, the degree of activation of a particular lexical item compared to its competitors has been found to vary as a function of factors related to the structure of the lexicon such as neighbourhood density and lexical frequency (Luce & Pisoni, 1998; Norris, 1992; McClelland & Elman, 1986). For example, words with higher usage frequency have been shown to reach a higher level of activation and therefore be accessed easier and faster compared to their less frequent competitors (e.g., Magnuson, Dixon, Tanenhaus & Aslin, 2007). In cases where there is a potential for lexical ambiguity due to reduced acoustic information (e.g., as a result of co-articulation or a phonological process) the structure of the lexicon might be even more relevant in accessing the intended word. In this chapter, the main goal was to investigate the incidence of minimal pairs ending in /t/-/p/ or /n/-/m/ in English lexicon, which can potentially become confusing as a result of place assimilation when occurring in certain phonological contexts. Even though the general frequency of assimilation/reduction in speech production has been previously reported in corpus data analysis (e.g., Dille and Pitt, 2007 on English data; Zimmerer, 2009 on German data), there are no studies that have specifically examined how the English lexicon is laid out with respect to words that can become lexically ambiguous as a result of place assimilation.

The differences in the correct recognition of words ending in assimilated or unassimilated nasals versus stops in Experiments 1-4 raises the question whether there are corresponding differences in the way the English lexicon is structured. Based on the claim that the acoustic cues to the place of articulation of nasals are not as salient as the cues to place of articulation of stops and therefore can potentially result in more lexical ambiguity, one might predict that minimal pairs ending in /n/-/m/ are less favourable to efficient language processing and therefore less likely to occur in the lexicon than pairs ending in /t/-/p/ (e.g., Graff, 2012). On the other hand, because similar sounding lexical items might differ in how frequently they are being used and in what contexts (e.g., phonological, semantic, syntactic, etc.), genuine lexical ambiguity might not be prevalent in real language use and therefore an asymmetry in the lexicon would not be expected.

Similarly, it is also possible that the lexicon is not shaped by considerations of perceptibility. In this case, there would again be no expectation to observe any asymmetry.

To test these predictions, the number of English minimal pairs that differ only in the place of articulation of their final consonant (/n/-/m/ or /t/-/p/) was calculated.<sup>13</sup> In addition, the average difference in frequency between the two words in each pair is reported and compared across the group of words ending in nasal versus stop consonants. Finally, cosines from Latent Semantic Analysis (LSA index; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) are reported for each minimal pair. These values range from -1 to 1, with higher values reflecting higher semantic relatedness between the two members of the pair in a given semantic context. The LSA values can serve as a proxy measure reflecting the potential for confusion that results from the ability to use both words in a given context.

## 5.1 English Corpus Analysis

The corpus analysis discussed in the current section is based on a search conducted using the online interface of the CELEX English lexical database (Relex V 0.4.4; Baayen, Pierpenbrock & Van Rijen, 1993). This database was consulted to establish a comprehensive list of minimal pair content words that differ only in the place of articulation of their final nasal or stop consonant (e.g., *cat-cap*, *line-lime*). From the original search within the online database, 152 minimal pairs ending in /n/-/m/ (with total 173,722 occurrences) and 174 minimal pairs (with 75,752 number of occurrences) ending in /t/-/p/ were identified. Because the CELEX corpus includes British rather than American English pronunciation and also includes words not in common use (e.g., *oaken-oakum*), the initial results were tailored accordingly such that the final list included only minimal pairs that were based on an American English pronunciation and no archaic or otherwise uncommon words. Finally, homophones were listed as distinct entries (e.g., *grate-grape*, *great-grape*) so that the usage properties associated with each word could be reported separately. As a result, a total number of 61 word pairs ending in nasal consonants and 89 word pairs ending in stop consonants were included in the final word lists. A Chi-Square test

---

<sup>13</sup> The minimal pairs did not include words ending in velar nasal or velar oral stop consonants. This is because the focus of this thesis is specifically on coronal-to-labial place assimilation and therefore patterns of confusability of word pairs ending in coronal vs. velar consonants are not available for comparison.

on the number of minimal pairs that were identified for each sound group indicates that the number of minimal pairs ending in /n/-/m/ is significantly fewer than pairs ending in /t/-/p/ ( $\chi^2(1, N = 150) = 5.22, p = 0.02$ ).

**Table 5-1.** Overall incidence, mean difference in lexical frequency, and mean semantic similarity index (LSA) for minimal pairs ending in /n/-/m/ or /t/-/p/ in English.

Minimal pair type	N	Freq. Diff (SD)	LSA (SD)
/n/-/m/	61	154.7 (302.7)	0.09 (0.12)
/t/-/p/	89	58.4 (122.1)	0.09 (0.09)

The lexical usage frequency of each word in the two word lists was obtained consulting the SUBTLEXus database of spoken American English (Brysbaert & New, 2009). For words that can be associated with different lexical categories, word frequency was acquired for the predominant part of speech. The absolute difference value between word frequencies for the words in each pair was calculated to arrive at an approximate measure of the relative usage frequency within each pair (Table 5-1). A small frequency difference would indicate the members of the minimal pair are closely similar in terms of how frequently they might appear in language. A larger difference in frequency, however, suggests that one member of the pair is much more likely to be used by speakers and therefore might be the preferred choice by listeners in case of perceptual ambiguity. Two pairs from /n/-/m/ list (*can-cam*, *man-ma'am*) and one pair from the /t/-/p/ list (*right-ripe*) were removed from the final analysis due to the unusually large difference (> 4000) between the frequency counts of the words in each pair. The results suggest that, on average, one member of the pair is more frequent than the other member for both nasal and oral stops. One sample t-tests against chance for the frequency difference data in each word list show that this difference is significant (i.e., different from zero) for both nasals ( $t(59) = 2, p < 0.001$ ) and oral stops ( $t(88) = 1.99, p < 0.001$ ). In addition, the average frequency difference between the members of the minimal pairs ending in /n/-/m/ is greater than the average frequency difference in minimal pairs ending in /t/-/p/ (155 vs. 58 respectively). This suggests



that, when the higher frequency member of the pair is intended by the speaker, there will be less lexical competition in the former group compared to the latter. Additionally, when the direction of this difference was also taken into consideration, it became clear that in the nasals list, the usage frequency of the words ending in a labial nasal is on average higher than the coronal nasals (133 vs. 37). However, in the stops list this direction is reversed with the coronal forms being on average more frequent than the labial forms (20 vs. 56). A two-sample t-test across the two word lists shows that in fact the average difference in frequency between the pair members is significantly higher for nasal list than for the stops list ( $t(145) = 0.28, p = 0.008$ ).

Latent Semantic Analysis (LSA) measures were retrieved using the online LSA interface (<http://lsa.colorado.edu>). As noted above, the LSA measure can provide a proxy measure of the potential for confusion resulting from a lack of semantic disambiguation. From the list of word pairs, the same outliers that were removed from the frequency analysis were also excluded for the current calculations. In addition, five words from the nasals list and two words from the stops list were not found in the LSA database; therefore, a total of 54 word pairs ending in nasals and 83 word pairs ending in stop consonants were included in the final analysis. Based on this number of tokens, the average LSA cosine was 0.09 for both nasal and stop word lists. The relatively small LSA index value that was found for both sound groups suggests that the words within each minimal pair from either of the sound categories do not tend to occur within the same semantic context. This suggests there is a low chance of semantic confusion between cases such as *line-lime* and *cat-cap*, when produced in natural language. A two-sample t-test that was conducted on the average LSA indices from the two lists did not show a significant difference between the two groups ( $p = 0.79$ ).

In sum, these results indicate that in fact the number of cases where two lexical items differ merely based on the place of articulation of their final nasal consonant is more limited compared to the number of the similar cases where the word final consonant is a stop. Moreover, it is also the case that on average the difference in lexical frequency between the members of minimal pairs within the /n/-/m/ group is greater than the difference between the members in the /t/-/p/ group making the former group of minimal pairs less confusable in case of perceptual ambiguity. However, in terms of the LSA measures, no significant difference was found between the two lists: The results showed a relatively small likelihood that the minimal pairs in either group tend

to co-occur in a given semantic context suggesting that the word pairs are less likely to be semantically related.

## 5.2 Discussion

The main aim of the current corpus analysis was to provide a better understanding of the structure of the English lexicon with regard to those words that have a potential for lexical ambiguity as a result of undergoing coronal-to-labial place assimilation. For this purpose, the measures included the number of minimal pairs, lexical frequency differences, and measures of semantic relatedness between the words in each pair. The results were contrasted for words ending in nasal versus oral stop consonants.

The results of the analyses revealed that the English lexicon shows some signs of being structured in a way that helps with tackling the issue of greater confusability of place of articulation in words ending in nasals versus those that end in oral stops. This is evident in the smaller number of minimal pairs ending in nasals (*line-lime*) compared to those ending in stops (*cat-cap*). This is in line with a study conducted by Graff (2012) on English lexicon in which he reported an inverse relationship between the degree of perceptual confusability of the sounds (based on confusability matrices from Miller and Nicely, 1955), and the number of minimal pairs containing those sound contrasts. Even patterns of diachronic merging of phoneme contrasts, for which the higher confusability of sounds has been shown to be an influential factor, are found to be related to the number of minimal pairs: The number of (lemma) minimal pairs with a particular phonemic contrast in a lexicon is negatively correlated with the probability of merging of that contrast (Wedel, 2013). Additionally, in the current data, a relatively larger frequency difference was found between the members of the /n/-/m/ minimal pairs compared to the words that belonged to the /t/-/p/ group. When there is a high potential for perceptual ambiguity, one of the two words in the former group is much more likely to occur and therefore would be expected as a more probable lexical candidate, reducing the overall confusion in spoken word recognition. A large body of research has in fact shown this effect of lexical frequency on recognition of words, where lexical items with higher usage frequency are generally activated and accessed faster (see Dahan & Gaskell, 2007; Dahan, Magnuson, Tanenhaus & Hogan, 2001 for examples using the eye tracking methodology employed here). Relatively higher frequency differences between the members of minimal pairs could be one of the factors in maintaining the word pairs

containing confusing sound contrasts (e.g., coronal versus labial nasal stops) in the lexicon (as opposed to merging the contrast, for example). In comparison, for minimal pairs ending in /t/-/p/, the frequency bias was found to be smaller. Therefore, upon hearing a perceptually ambiguous form from this latter group, both words would be expected to compete strongly as lexical candidates on grounds of frequency. However, the relatively more distinct sound contrasts in these word pairs could help steer listeners toward the intended candidate.

These findings suggest that, overall, the chance of encountering a lexical item being ambiguous between ending in /n/ or /m/ is much smaller than a lexical item ending in an ambiguous /t/ or /p/. This in turn may mean that the pattern of the lexicon (likely resulting from gradual historical change rather than conscious choices of altruistic speakers) complements the patterns of both production and perception of assimilated forms ending in nasal or oral stops. Even though place assimilation is more frequent in words ending in coronal nasals (despite the fact that place cues are perceptually less salient in this sound category), the efficiency of the recognition system can still be maintained through the way that the lexicon is structured.

## Chapter 6 Conclusions

Successful language communication is often achieved despite considerable variation in the speech signal that results from connected speech processes. In the recognition of phonologically assimilated words, the acoustic details corresponding to the place of articulation of an assimilated sound have been claimed to be an important source of information for correctly identifying underlying place of articulation of a given sound. Additionally, a large body of research has suggested that the viability of the phonological context (i.e., whether the following sound is one that triggers place assimilation) has an effect on compensating for place assimilation. The focus of the current thesis was on investigating the nature of the effect that phonological context has on compensation for assimilation, and whether and in what ways processing of assimilated forms is influenced by factors such as the type of speech sound in question and the structure of the lexicon. In the following section, I provide a summary of the experimental results reported earlier, and address the three core questions proposed in the first chapter of this thesis (listed below):

1. How does the recognition system make use of the available phonological context information?
2. How do the inherent differences in characteristics of the nasal and oral stops affect the correlation between the acoustic cues and phonological context information?
3. Does structure of the lexicon work against or in favor of resolving phonological variation?

I then describe the implications of the results with respect to the major accounts of how listeners deal with assimilation during language processing.

### 6.1 The Effect of Phonological Context on Perceptual Judgements

Previous research on the effect of the phonological context on compensation for assimilation has, for the most part, focused on the viability effect of the following phonological context. This work has provided some evidence supporting this effect (e.g., Gaskell & Marslen-Wilson, 2001; Gow & McMurray, 2007; Mitterer, Csepe & Blumert, 2006). The difference that was found between the recognition of assimilated words in viable and unviable contexts is generally

assumed to be due to the facilitatory effect of the triggering context on processing place assimilation. However, in the majority of the research conducted on this topic, the comparison case involved a condition where word forms were extracted from their natural context and spliced into a context that would or would not trigger assimilation. As noted in the introduction to Chapter 3, an issue with this comparison is that the observed difference could instead stem from the effects of hearing an unviable context after an assimilated sound (i.e., mismatching information resulting in perceptual confusion).

To provide a different test of the effect of the phonological context on the recognition of words with assimilated final consonants, two sets of experiments were conducted on naturally produced unassimilated and assimilated words. Also, unlike many previous studies on place assimilation, where the stimuli consist of a mixture of words ending in either nasal or oral stops, each sound category was addressed separately using a relatively large stimulus sample (24 word pairs each). Another important feature of the stimuli used in this research was that the assimilated version of the stimulus words was produced and presented in their natural phonological context rather than being cross-spliced into an unviable context, as is often the case. Although, edited stimuli were also used in the current study to prepare the isolated words, the difference was that the “trimmed” versions were incomplete rather than being accompanied by a context that would provide misleading information by not triggering regressive place assimilation. In fact, incomplete perceptual information is not uncommon in natural language use, where a phone call is cut off in the middle of a conversation or bad signal quality interrupts the news on the radio. This technique allowed a focus on how the availability (as opposed to viability) of the triggering context can affect the recognition system when processing assimilated forms. In the first set of experiments (Experiments 1 & 2), English listeners heard strongly/completely assimilated or unassimilated words ending in either nasal or oral stop consonants played in isolation (without the following phonological context being available). In the second set of the experiments (Experiments 3 & 4), listeners heard the same words, but in their original carrier sentence where the following phonological context was now available. In the first set of experiments the assimilated words across the two sound groups were most often perceived as ending in a labial rather than a coronal consonant (average percentage of coronal responses = 15%). However, in the second set of experiments, where the listeners could also hear the phonological context, the rate of the correct identification of assimilated words was improved by 26% on average. In

addition, a comparison between prime response times from the first set of the experiments and the second set of the experiments indicated lexical processing facilitation upon provision of the phonological context in the second set.

These findings provide clear evidence that the presence of the phonological context in fact affects the perceptual system during the processing of assimilation, even when assimilation is strong/near complete. This is consistent with Gaskell and Snoeren (2008, Experiment 2), who showed a viability effect of the phonological context even for canonically labial (or velar) consonants, regardless of the provision of a related semantic context, which in an earlier study was found to be necessary in the processing of completely assimilated words that are lexically ambiguous (e.g., *rum-run*, Gaskell & Marslen-Wilson, 2001). Crucially, the current results confirm the role of the post-assimilation context in biasing the perception of place of articulation of assimilated word-final consonants through a more direct comparison with a condition where this information is not provided.

## 6.2 Nasal versus Oral Stop Place Assimilation

Traditionally, place assimilation has been used as an umbrella term to cover those connected speech variations that presumably affect various consonants in a similar way in that their place of articulation becomes more similar to the place of articulation of the following (or preceding) consonant. However, analyses of probabilistic patterns and acoustic and perceptual characteristics of assimilated sounds, among other evidence, suggest that distinct sound groups, such as nasal versus stop consonants, are affected differently by place assimilation (e.g., Boersma, 1998; Dilley & Pitt, 2007; June, 1995; Steriade, 2001). This in turn indicates that there might be differences in the mechanisms involved in processing of assimilation that are dependent on the characteristics of the consonant that has been assimilated. Therefore, the research described in the previous chapters considered nasal and oral stops independently to investigate these possible differences.

An acoustic analysis of the experimental stimuli was conducted to evaluate and validate the unassimilated and assimilated word-final consonants in terms of their acoustic characteristics as well as degree of assimilation. The results of this initial analysis showed that whereas the duration of assimilated oral stop consonants was shorter than the duration of coronal and labial word-final stops (complying with the pattern shown in Dilley & Pitt, 2007) and other measured

acoustic cues that were based on F2 frequency showed intermediate tendencies, the measures for assimilated nasals were more similar to corresponding measures for labial rather than coronal nasal consonants. This result suggested that the degree of assimilation of the word-final nasal consonants in the current stimuli was perhaps acoustically stronger than the degree of assimilation in the word-final stop consonants.

The results of the identification tasks in the experiments matched the patterns observed in the acoustic analysis of the stimuli. When the assimilated words were heard in isolation, listeners frequently perceived the words as ending in a labial consonant. There was, however, a difference in identification rates between the words ending in nasals and the words ending in stop consonants. Specifically, listeners were slightly better at identifying the intended place of articulation of the final assimilated stops as coronal (21% correct), whereas they almost always misidentified the intended place of articulation of the assimilated final nasals (only 8% correct). Together with the results of the acoustic analysis, these findings suggest a strong degree of assimilation for the assimilated word-final oral stops and a (near) complete assimilation for the word-final nasal consonants in the current naturally produced stimuli. Additional comparisons of the results from Experiments 1 and 2 with the results from Experiments 3 and 4, where the assimilation triggering phonological context was also provided, suggested an overall improvement in perception of the assimilated forms as coronals in the second set of experiments. Interestingly, however, despite the differences in acoustic characteristics the degree of influence of the phonological context on listeners' final decision in the identification tasks was approximately the same between the words ending in assimilated nasals and the words ending in assimilated stop consonants (~ 26% increase in coronal responses). This indicates that the phonological context has an almost automatic and consistent positive impact on recognition of assimilated words regardless of the consonant type.

Further comparisons were made between the two groups of consonants based on the mouse-click response times for prime words in Experiments 1 and 2, which used isolated words ending in nasal or stop consonants respectively. The listeners were overall faster to correctly recognize words ending in stops compared to those ending in nasals even when the word-final sounds were unassimilated and therefore unambiguous. This supports a separate claim stating that cues to place of articulation in nasal consonants are perceptually less salient than in stops, resulting in greater indeterminacy and therefore longer RTs. In addition to the overall difference in RTs

between isolated words ending in nasals versus those ending in oral stops, RTs were found to be slower for assimilated prime words ending in nasals that had a minimal pair lexical competitor (e.g., *line* where *lime* is also a real word). The same effect, however, was not observed for prime words ending in assimilated oral stops. This indicates the indeterminacy that is reflected in the results involves some degree of lexical competition for the words ending in nasal stops, even though the minimal pair lexical competitor was not displayed.

The response time for prime sentences were also compared between Experiments 3 and 4 and with the results from Experiments 1 and 2. When the prime sentence RT results were compared between Experiments 3 and 4 (nasal vs. oral stops), the patterns suggested no difference in processing time for prime words ending in assimilated nasals compared to assimilated oral stops. Additionally, the RTs were no longer slower for words ending in nasals that had a minimal pair competitor compared to the words that had no minimal pair. These do not match with the patterns that were observed in Experiments 1 and 2, where the RTs were generally slower, especially for words ending in nasals that had a minimal pair. The contrast in patterns appears to reflect the facilitatory effect of the triggering context in processing assimilated words ending in nasals. This outcome supports the notion that there may be somewhat distinct mechanisms involved in processing assimilation for the two groups of consonants.

In addition to the identification and RT analyses, an analysis of eye movements as an online measure of spoken-word recognition was also conducted in Experiments 3 and 4. The results of this latter analysis revealed differences between the processing of the assimilated forms ending in the two types of consonants. Recall that a competitor item was always present on the display for the target word. As speech unfolded over time, the proportion of fixations to the target item relative to the other items on the display, particularly the competitor, was considered. When the final consonant of the target word was unassimilated, there was a clear advantage of the fixations to the target item over the competitor. For both groups of words that ended in unassimilated nasals or oral stops, this target advantage started at the offset of the consonant and within the timeframe where the acoustic cues corresponding to the following phonological context would become available. Crucially, the results further showed that a recent instance of “undoing” place assimilation might result in some degree of priming effect for target words ending in an assimilated coronal nasal (Experiment 3). This effect occurred at the point when the triggering phonological context following the subsequent target word would become available.



Specifically, when the target word was assimilated and the prime word was unassimilated, listeners were more likely to visually consider the competitor word (e.g., *lime*) compared to the intended target word (e.g., *line*) by the offset of the nasal consonant. This pattern reflects the results of the offline identification task where there was a clear bias in listeners' final selection in choosing words ending in the labial nasal upon hearing an assimilated word. When the previously-heard prime word was assimilated, however, this suppressed the consideration of the competitor over the target from the point where listeners heard the onset of the following labial context. Strong competition then continued between the target and competitor item as speech unfolded in time, until 600 ms later, which is around the time when the click responses were made. No such priming effect was obtained for stop place assimilation in words ending in assimilated stop consonants (Experiment 4). The priming effect observed in online data, however, was not reflected in the offline identification results. This contrast in the measures is not unexpected considering the high sensitivity of eye movements to the temporal details of processing as speech unfolds in real time. Traditional offline measures (which are based on a conscious decision), such as forced choice identification, have been shown to fail to capture subtle biases that might affect processing before the final decision has been made (e.g., Ben-David, Chambers, Daneman, Pichora-Fuller, Reingold & Schneider, 2011). Therefore, despite the comparable effect of the phonological context on listeners' final perceptual judgement, the fine-grained differences in processing of nasal and oral stop place assimilation again suggest that the processing of these otherwise similar variations might not be the same.

Mitterer (2011) proposed different mechanisms for processing of place assimilation/reduction in Dutch nasal and stop consonants. He did not find an effect of contextual viability for [t]-reduction in a discrimination task and in an identification task in an earlier study (Mitterer & McQueen, 2009) but found a viability effect in an eye tracking study and concluded that the compensation for [t]-reduction is only triggered at later stages of processing. The current identification results, however, clearly show an improvement upon provision of the following phonological context for both assimilated nasal and oral stops (in English) even in the identification task. In addition, Mitterer's study only focused on a very small number of stimuli (only one pair of pseudo-words was tested for each sound) and the difference in strength level of assimilation between the word-final nasal and oral stop consonants was a major concern in interpretation of the results of the study (very strong [t]-reduction versus weak nasal place

assimilation). As was supported by the acoustic and perceptual measures, both nasal and stop consonants in the stimuli used here were naturally produced with strong/complete assimilation, using a relatively large list of real words (24 set of prime and target pairs for each sound group), making the comparisons between nasals and oral stops more valid.

The patterns of prime RTs and the priming effect observed for nasals in the current experiments suggest a dissociation in the recognition of words ending in assimilated nasal and oral stops that does not seem to be fully explained based on the degree of assimilation of these sounds. In addition to an overall higher processing load for isolated words ending in nasals, the effect of the triggering context for assimilation on the process of undoing assimilation also seems to differ. Even though the results of the offline identification task suggest a similar amount of contextual facilitation in recognition of both forms, online data (RTs and eye movements) shows a divergence between the two groups. These results are consistent with the idea that there may be differences in mechanisms for integrating the phonological context information in processing of different classes of assimilated forms: Specifically, although contextual information has a clear influence on explicit judgements of the word identity for both words ending in assimilated nasals and assimilated oral stops, an earlier sensitivity to this information was observed only with words ending in assimilated nasals. This is broadly consistent with the claim that was made by Mitterer (2011) about assimilation processes in Dutch.

A question that might be raised is whether any potential variation in rate of speech production between the instructions from the two sound groups might have influenced the perceptual effects. Assimilated forms are more often perceived as such with faster speech rates (Li & Kaiser, 2012), presumably because faster speech often results in more assimilation (and other forms of variation related to connected speech). If the rate of speech was significantly different between the utterances carrying assimilated or unassimilated words in either prime or target sentences, this could have influenced listeners' decisions. To test for any possible variation, rate of speech was measured for both nasal and stop prime and target carrier sentences. Because in naturally produced connected speech, assimilation can affect consonant duration and the pause between the assimilated and the following assimilating consonant, production rate was calculated for the segments immediately before the word-final nasal or oral stops rather than the whole carrier sentence. Therefore, rate of speech was calculated based on the number of segments produced preceding the word-final consonant per second (sps). For both prime and target instructions, the

measures were taken from the onset of the word *click* for the purpose of consistency. For example, the rate of speech for the utterance carrying the prime word *bean* was 13.2 sps, calculated by dividing the number of segments preceding [n] (i.e., ten segments in this example) by the amount of time from the onset of the word *click* to the onset of [n] in *bean* in seconds (0.758 s).

Based on the results, rate of speech for prime or target carrying utterances did not show much variation across assimilated and unassimilated conditions for the nasal group (M = 13.5 sps and M = 13.4 sps respectively). However, rate of speech was slightly faster for the target sentences compared to the prime sentences (M = 12.6 sps and M = 14.2 sps respectively). For the oral stop group, in addition to the lower rate of speech in prime sentences compared to target sentences (M = 12.7 sps and M = 14.8 sps respectively), assimilated forms were also on average slightly faster than unassimilated forms in prime sentences (M = 13.5 sps and M = 12.3 sps respectively). Two separate between-item two-way repeated measure analysis of variance (ANOVA), with Assimilation (unassimilated vs. assimilated) and Sentence Type (prime vs. target) as factors, were conducted on the average rate of speech for nasal and oral stop groups. For both nasal and oral stops there was a highly significant main effect of Sentence Type ( $F(1, 92) = 71.1, p < 0.001$  and  $F(1, 92) = 122, p < 0.001$  respectively) indicating that target sentences were overall faster than prime sentences. Additionally, for oral stops, there was a significant main effect of Assimilation ( $F(1, 92) = 7.4, p = 0.007$ ). No other significant main effect or interaction was found in either group (all  $p$ 's  $> 0.42$ ). The difference found between the rate of segment production in the prime compared to target utterances in both groups is not surprising considering that longer utterances are generally found to be produced faster compared to shorter utterances (e.g., Haselager, Slis & Rietveld, 2009; Malecot, Kizziar & Johnston, 1972); and in the current case, the overall length of the target sentences was greater than the prime sentences due to the additional word *now* and the pause between this word and the word *click* that was present only in target sentences. The difference between rate of speech of the assimilated and unassimilated prime sentences in oral stop group is, however, rather unexpected, especially because a similar pattern was not observed in the target sentences. One might argue that the difference in rate of speech of the assimilated and unassimilated prime sentences in the oral stop group might have had an effect on priming results. Because reduction in general, and place assimilation in particular, is more probable in faster speech, one could equally argue that the

faster speech rate should have boosted the expectation of place assimilation, in turn providing a stronger priming cue for phonological processing in the oral stop condition. However, this was not the observed pattern of the results in the current experiments, where an effect of priming was observed only for nasals.

### 6.3 Structure of the Lexicon and Place Assimilation

In the English corpus analysis reported in Chapter 5, I explored lexical characteristics of minimal pair words ending in nasal consonants /n/-/m/ versus pairs ending in stop consonants /t/-/p/. These are words such as *line* or *cat* (similar to the stimuli used in Experiments 1-4), where the assimilation process can result in lexical confusability due to the perceptual similarity of the assimilated words to words such as *lime* and *cap* respectively. Many models of spoken word recognition predict an increase in the amount of lexical competition for the lexical candidates that are perceptually similar (e.g., NAM, Luce & Pisoni, 1998; TRACE, McClelland & Elman, 1986). The results of the current experiments (see Chapters 3 and 4 on comparison of the RTs for ambiguous vs. unambiguous prime words) as well as previous research (e.g., see Gaskell & Marslen-Wilson, 2002) support this claim by showing the relatively higher mental processing demand to resolve such lexical confusions. The current corpus analysis investigated the prevalence of these minimal pairs, and the extent to which differences in their usage frequency and semantic similarity (LSA index) would be correlated with the potential for confusion.

Results demonstrated that compared to potentially confusing minimal pairs ending in a coronal or labial oral stop consonant, there is a significantly lower number of word pairs ending in a coronal or labial nasal consonant. This suggests that even though assimilated coronal nasals can be perceptually more confusing in terms of their place of articulation, there is a lower chance that this perceptual confusion would lead to lexical ambiguity in natural language use. Another interesting finding was that within the minimally contrastive word pairs that were found for both groups, those ending in a nasal were more distinct from each other in terms of usage frequency compared to those word pairs that ended in a stop consonant. The relevance in this case is that lexical frequency has been shown to have an important effect on lexical access (alongside neighborhood density) in that the higher the relative usage frequency of a lexical item, the faster and the more efficiently it can be accessed (Luce, 1986; Segui, Mehler, Frauenfelder & Morton, 1982). Thus, when the higher frequency member of a minimal pair is used, a larger frequency

difference in relation to the low-frequency alternative would mean that less competition would result. The fact that the frequency gap between high and low frequency members of a minimal pair are greater for nasals could therefore be understood to mean that word recognition might be able to use this information to compensate for higher confusability of this sound group, at least when the intended word is the higher-frequency member of the pair. Finally, the possibility of both members of minimal pairs ending in nasal or oral stops to occur in a similar semantic context is very low. This would mean that in cases where place assimilation can potentially result in lexical confusion, higher-level information such as semantic context is likely to help listeners overcome confusion.

These results suggest that the structure of the lexicon is in line with the patterns that are observed in production and perception of assimilation in English words. In comparison with lexical items that end in coronal stop consonants, the recognition of lexical items that end in coronal nasal consonants can be especially challenging. On the one hand, a number of previous studies have demonstrated that the place contrasts in nasal consonants are generally less salient than in oral stop consonants, which in turn results in relatively higher confusability of place of articulation in nasals compared to oral stops (e.g., June, 2004; Kawahara & Garvey, 2014; Ohala & Ohala, 1993). This claim was supported by the identification results in the current study, where listeners' perception of the place of articulation of the coronal nasal consonants was relatively poor compared to their perception of the place of articulation of the oral stops, even when the consonants were in their canonical form. On the other hand, previous corpus analyses have shown that place assimilation is more frequent in words ending in coronal nasals compared to the words ending in coronal stops (e.g., Dilley & Pitt, 2007). In fact, this is in accordance with the proposal that confusable contrasts more often undergo reduction (e.g., Kawahara & Garvey, 2004).

Considering the asymmetries that exist between assimilated nasal and oral stop consonants, both in production and perception, it may be that the lexicon has evolved so that it partially compensates for the higher perceptual confusion of the minimal pairs ending in nasals. This ultimately renders the recognition system more efficient for handling acoustic variation. Although this proposal is speculative, it is in line with the notion of "effective contrast" proposed by Ussishkin & Wedel (2002, 2009), according to which instead of only maintaining sound-level contrasts, the language system is based on maintaining a type of functional contrast that is

influenced by a combination of the factors including lexical statistics such as lexical frequency. Under conditions where the frequency contrast is large, this account would predict that the sound contrast could be relatively weak as long as the overall “effective contrast” compensates for it.

## 6.4 Implications for the Accounts of Compensation for Assimilation

Although testing the full range of predictions put forward by existing theories was not the main goal of this thesis, the findings have implications for contemporary accounts of the perceptual processing of assimilated consonants. As discussed in Chapter 1, the three major processing accounts involve the notions of feature parsing, phonological inference and perceptual integration. Also, featural underspecification has been proposed as a representational account for the place assimilation process.

The underspecified representation account, which was later modified into the featurally underspecified lexicon (FUL) model, suggests that in a language like English, phonemes such as /n/, /t/ or /d/ are underspecified for the place of articulation (i.e., the feature [coronal]) at the level of lexical representation (Lahiri & Reetz, 2002, 2010; Wheeldon & Waksler, 2004). According to this account, acoustic cues to labial and velar place of articulation match with the specified [labial] and [velar] features respectively but at the same time would not mismatch with the underspecified [coronal] feature. On the other hand, acoustic cues to coronal place of articulation would not match with the [labial] or [velar] features and would only match with [coronal] feature. Therefore, based on FUL, a word such as *lime* that ends in a labial consonant would be perceptually ambiguous between *lime* and *line*, whereas *line* would be unambiguous. The role of the contextual information (phonological, semantic, etc.) is assumed to be secondary, if at all influential. Similarly, on this account, the role of residual acoustic place of articulation cues in resolving assimilation is minimized because accurate perception of assimilated sounds should be achieved at the level of featural representation and through underspecified versus specified distinction among sound features, as was discussed earlier. Therefore, the occurrence of strong/complete place assimilation in the stimuli, as was the case in the current research, is not unexpected in this account. The current results do not completely support the predictions of this account, however. With, on average, 15% coronal responses (only 8% for nasals) for strongly/completely assimilated words in isolation, a coronal interpretation does not seem to have been a likely option. Since in the recognition of these isolated words, no phonological

context was available to either support or refute the original extraction of the features, the underspecification account would predict the assimilated forms to be equally mapped onto either a labial or an underspecified coronal lexical representation, leading to something closer to 50% response. More importantly, the facilitatory effect of the triggering phonological context on perception of the assimilated forms as soon as the context was made available to the listeners (Experiments 3 & 4) would again be unexpected based on the underspecification account, which does not assume a crucial role for the phonological context effect.

Accounts that propose an active role for phonological context can better explain the patterns observed in the current study, namely the misidentification of assimilated words in isolation and the clear effect of the phonological context on biasing listeners' perception. The three accounts of feature parsing, phonological inference and perceptual integration all assume a role for bottom-up information in the input signal relating to the segment in question, as well as the triggering assimilation context that follows this segment. For the feature parsing and perceptual integration accounts, the role of the context is more prominent at the lower auditory level. The feature parsing account predicts an effect of phonological context only for partially assimilated segments that have retained some residual acoustic cues to the underlying place of articulation (Gow, 2002; Gow & Im, 2004; Gow & McMurray, 2007). When a partially assimilated consonant is followed by a viable phonological context for assimilation, the feature parsing account predicts no lexical ambiguity as a result of place assimilation. This is because the process of feature parsing is expected to assign the perceived acoustic cues corresponding to the labial feature to the following labial consonant and the remaining coronal cues should be associated with the underlying place of articulation of the assimilated consonant, resolving any perceptual ambiguity. If the assimilated consonant does not carry any residual cue to the coronal place of articulation, as is the case with complete assimilation, again the feature parsing account does not predict any perceptual ambiguity. In this latter case, the prediction is that the listeners should take the labial cues at face value and perceive the assimilated sound as a labial sound. The feature parsing account would therefore predict the assimilated oral stops in Experiment 4 (with context being present) to be perceived as having a coronal place of articulation because according to the acoustic and perceptual measures, these stops have undergone a near-complete (rather than complete) place assimilation. Contrary to this prediction, however, the provision of the triggering phonological context in this experiment led to lexical ambiguity, where listeners

identified assimilated forms as labial stops 50% of the time. The patterns observed in perception of the completely assimilated nasals in Experiment 3 are also in conflict with the predictions of the feature parsing account. Feature parsing does not account for compensation for completely assimilated forms, since in such cases there would be no residual acoustic cue that can be attributed to the coronal feature of the underlying place of articulation of the assimilated consonant. However, the provision of the phonological context in Experiment 3 clearly resulted in an increase in coronal responses. These results imply that (partial) compensation for assimilation takes place even when the assimilation is strong/complete even though the end result can still involve lexical ambiguity in some cases.

The phonological inference and perceptual integration accounts, on the other hand, do not assume a crucial role for presence of residual acoustic cues to the underlying place of articulation in compensation for assimilation (e.g., Mitterer, Csepe & Blomert, 2006; Gaskell & Marslen-Wilson, 1996). The context effect observed in the current set of experiments can be explained more comprehensively within these frameworks, even given the case of the more complete assimilation that was observed in the nasal stimuli. According to the perceptual integration account, the general auditory system has a tendency to maintain the contrastive properties in acoustic signals. This in turn can lead to perception of an assimilated segment and the triggering context as perceptually distinct sounds even when strong assimilation results in their acoustic similarity. Based on the inference account, however, probabilistic information and listeners' phonological knowledge has a more prominent role in recovering the underlying form of even extremely reduced sounds. Both accounts, therefore, accurately predict the bias that was observed towards interpreting an assimilated consonant as a labial segment when there was no context available.

Nevertheless, the distinction in processing mechanisms that was observed between nasal and stop consonants based on the prime response times and eye movement priming results can be explained more easily within an inference framework than the perceptual integration or any framework that is solely based on lower-level auditory processing mechanisms. The mechanisms underlying the general auditory processing are considered to be automatic and affecting earlier stages of processing of speech signals. As a result, in frameworks that follow the general auditory approach in processing place assimilation (i.e., feature parsing and perceptual integration accounts), the acoustic cues of sounds, such as formant frequencies, and the acoustic



information corresponding to the following sound (i.e., the phonological context) are unconsciously and automatically processed at an early stage through the general auditory system and the type of acoustic signal (e.g., speech or non-speech) is not expected to have an influence on this process (e.g., Viswanathan, Magnuson & Fowler, 2010). Therefore, an account such as perceptual integration would predict the effect of the phonological context on perceptual processing of assimilated nasals and assimilated oral stops to be similar. On the other hand, the patterns observed in the prime RTs in the current study indicate that the processes that are involved in compensation for nasal place assimilation might not be the same as those in stop place assimilation. Specifically, in the processing of the prime words that had a minimal pair/potential competitor, a relatively longer processing time was only observed for isolated words ending in assimilated nasals and not for words ending in assimilated oral stops. But perhaps more importantly, while the provision of the following phonological context facilitated processing for both groups of words, it further reduced the processing load for words with a minimal pair that ended in a nasal consonant suggesting that the amount of lexical competition was decreased only in this group of words.

Further evidence on this point comes from the dissociation that was observed based on the results from priming eye fixations: The phonological context information integrated at early stages of lexical access for words ending in assimilated nasals whereas this information seemed to mainly influence the final decision making stage for assimilated oral stop consonants. Such a dissociation can be explained within the inference framework considering the probabilistic and distributional differences that are found to exist between the two groups of sounds in previous corpus analyses (see Dilley and Pitt, 2007), which was also reflected by the results of the corpus analysis (Chapter 5). Since in English, word-final coronal nasal consonants are often heard as assimilated to the place of articulation of the following labial (or velar) consonant, listeners use this probabilistic information when they hear a word-final nasal that carries acoustic cues to labial place of articulation and integrate the phonological context information right at the moment it becomes available to modify their perception of the assimilated nasal sound. For oral stops, however, the effect of the phonological context seems to be evident only at the decision stage, which can explain the lack of an early priming effect when the assimilation triggering sound is heard. Neither the feature parsing account nor the perceptual integration account predicts such differences in processing similar phonological variations that affect speech sounds.

The current findings therefore most closely match with the predictions of the phonological inference account. The overall facilitatory effect of the assimilation context on perceptual processing of the assimilated forms and the evidence suggesting a dissociation in the mechanisms involved for integration of the acoustic and contextual information in processing of assimilated nasals versus stop consonants can both be successfully explained based on the phonological inference account. Nevertheless, the low percentage of coronal identification for strongly/completely assimilated words that were used in the current study is a strong indication of the important role that bottom-up information plays in resolution of perceptual ambiguity in the absence of semantic or other higher-level contrastive information. Therefore, the current findings are in line with an account that would combine auditory mechanisms and statistical learning mechanisms for processing place assimilation.

## 6.5 Concluding Remarks

The recognition of spoken assimilated words, especially when there is a potential for lexical ambiguity, is affected by phonological context information. The current findings have provided direct evidence to support the claim that this sensitivity is rather robust and independent from factors such as degree of assimilation (i.e., partial versus strong/complete assimilation) or the type of sound that is affected (i.e., nasal vs. oral stop consonants). Nevertheless, phonological context information was found to be insufficient for recovering the underlying form in the absence of strong acoustic cues to coronal place of articulation or higher level information. This was the case despite the fact that the triggering phonological context in the current experiments was always predictable (the word *button*), eliminating the possibility of misperception of the following context. The findings have also provided evidence that, in contrast to the typical assumption, the recognition system does not necessarily handle the outcome of the assimilatory processes affecting different sound classes the same way. In particular, the results of the online measures of processing show that even though the phonological context information has an overall comparable facilitatory influence on the recognition of words ending in assimilated nasals and oral stops, the underlying processing mechanisms might vary depending on the characteristics of the sound in question. Moreover, the difference observed between response times for prime words ending in nasal versus oral stops, even when these consonants were unassimilated, provides additional evidence supporting the hypothesis that place contrasts in nasal stops are overall less salient than in oral stops (e.g., Kawahara & Garvey, 2014). Finally,

the current findings provide evidence supporting the claim that the structure of the lexicon might in fact be in accordance with the patterns of production and perception data of regular alternations such as place assimilation. Specifically, the prevalence of certain minimal pairs, combined with patterns of use, may help compensate for the greater tendency for nasals to entail lexical ambiguity due to assimilation and the salience of place contrasts.

In investigating the factors influencing the process of compensation for place assimilation, the current work has focused on studying naturally produced stimuli. This choice was made due to the fact that most previous studies only focused on assimilated forms that were substantially manipulated in various aspects, and studying naturally produced speech provides the much needed information regarding processing place assimilation as it happens in the real language comprehension. Nevertheless, using naturally produced speech would mean some factors could no longer be completely controlled. Among these factors are rate of speech and degree of assimilation (i.e., weak, strong or complete). As discussed in earlier sections, in this work an effort was made to account for the potential influence of these two factors. In future studies, the potential effects of these factors can be tested using similar stimuli that are normalized for speech rate and degree of assimilation. The latter factor is more challenging because different sound categories involve features that are not necessarily comparable. Testing listeners' recognition of words that are produced with various degrees of assimilation across a continuum is one possible way for controlling the effects of this factor more directly. Additionally, similar to the majority of previous research on this topic, the focus of the comparisons in the current research was on coronal-to-labial place assimilation, and only voiceless oral stops were included. Voiceless oral stops were chosen considering the results of earlier corpus analyses that indicate voiced stops undergo place assimilation less often than either nasal or voiceless oral stops (Dilley and Pitt, 2007). For further generalization of the current findings, coronal-to-velars place assimilation and assimilation in voiced stops could be tested using a similar methodology for comparison purposes. There are also directions for future work on analysis of the structure of the lexicon. Due to limitations on time, the present study exclusively focused on exploring the lexicon with respect to the type of the contrast used in the current experiments. Expanding the corpus data analyses to other minimal pairs that do not undergo place assimilation, or pairs that can be affected by phonological processes other than place assimilation, would be beneficial.

In sum, the results of this thesis contribute to our general understanding of how listeners achieve lexical access in connected speech, and highlight important factors that are not fully addressed in competing theories of compensation for phonological variation. To diagnose the underlying mechanisms involved in the process, the issue was explored from a perspective that emphasized often-overlooked differences in place assimilation in nasal versus oral stops and its effects on perceptual processing. Also, an eye tracking methodology combined with a (process) priming paradigm provided insights into the real-time processing of variation in naturally produced speech and how it can differ from listeners' final perceptual judgements. From specific characteristics of individual speech sounds to the general patterning of the lexicon, it has been shown that a variety of factors need to be considered to understand how listeners ultimately arrive at the accurate recognition of spoken words. These factors modulate integration of various sources of information at each level of processing, reflecting the dynamic nature of language comprehension phenomena.

## References

---

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38(4), 419-439.
- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119(5), 3048-3058.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bateman, N. (2011). On the typology of palatalization. *Language and Linguistics Compass*, 5(8), 588-602.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4), 2350-2366.
- Ben-David B, Chambers C, Daneman M, Pichora-Fuller M, Reingold E, Schneider B. (2011) Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, 54(1), 243–262.
- Bien, H., & Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: evidence from explicit categorization and the N100. *Frontiers in Psychology*, 4, 21.
- Boersma, P. (1998) *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics
- Boersma, P. & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.23, retrieved 15 September 2013 from <http://www.praat.org/>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera 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.
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

- Caramazza, A. & Yeni-Komshian, G. (1974). Voice Onset Time in Two French Dialects. *Journal of Phonetics*, 2, 239-245.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Clayards, M., Niebuhr, O., & Gaskell, M. G. (2015). The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. *Attention, Perception, & Psychophysics*, 77(1), 311-328.
- Coenen, E., Zwitserlood, P., & Bolte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language & Cognitive Processes*, 16(5-6), 535–564.
- Coleman, J. (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31(3), 351-372.
- Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, 23(3), 235-245.
- Cutler, A. (1995). Spoken word recognition and production. In J.L. Miller & P.D. Eimas (Eds), *Speech, language, and communication*, (pp. 97-136). New York: Academic Press.
- Cutler, A. (1995). Spoken-word recognition. In G. Bloothoof, V. Hazan, D. Hubert, & J. Llisterra (Eds.), *European studies in phonetics and speech communication*, pp. 66-71. Utrecht: OTS.
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations?. *Cognition*, 108(3), 710-718.
- Dahan, D., & Gaskell, G. M. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57(4), 483-501.
- Dahan, D., Magnuson, J., & Tanenhaus, M. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42(4), 317-367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5-6), 507-534.
- Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36(3), 704.
- Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. *Variation and gradience in phonetics and phonology*, 14, 265.

- Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. *Laboratory Phonology*, 9, 411-442.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391.
- Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America*, 122(4), 2340-2353.
- Dinnsen, D., & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, 12(1), 49-60.
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Elman, J., & McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27(2), 143-165.
- Ernestus, M. (2003). The Role of Phonology and Phonetics in Dutch Voice Assimilation. In: J. van de Weijer, V. J. van Heuven, & H. van der Hulst (eds.), *The Phonological Spectrum, Volume I: Segmental structure*, pages 119-144. Amsterdam: John Benjamins.
- Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes*, 21(7-8), 1011-1029.
- Gaskell, M. G. (1994). *Spoken word recognition: A combined computational and experimental approach*. University of London.
- Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31(3), 447-463.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1), 144-158.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 380-396.
- Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 1632-1647.
- Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1211-1238.

- Goldsmith, J. A., Riggle, J., & Alan, C. L. (Eds.). (2011). *The handbook of phonological theory* (Vol. 75). John Wiley & Sons.
- Gordon-Salant, S., & Fitzgibbons, P. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments, *The Journal of the Acoustical Society of America*, 119(4), 2455–2466.
- Gordon-Salant, S., Yeni-Komshian, G., & Fitzgibbons, P. (2008). The role of temporal cues in word identification by younger and older adults: Effects of sentence context. *The Journal of the Acoustical Society of America*, 124(5), 3249.
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45(1), 133–159.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163–179.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Attention, Perception, & Psychophysics*, 65(4), 575–590.
- Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51(2), 279–296.
- Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation, *Papers in Laboratory Phonology*, 9, 173–200.
- Gow, D. W., & Segawa, J. A. (2009). Articulatory mediation of speech perception: a causal analysis of multi-modal imaging data. *Cognition*, 110(2), 222–236.
- Graff, P. (2012). *Communicative Efficiency in the Lexicon*. Doctoral dissertation, Massachusetts Institute of Technology.
- Hale, M., Kissonock, M., & Reiss, C. (2007). Microvariation, variation, and the features of universal grammar. *Lingua*, 117(4), 645–665.
- Halle, P., & Adda-Decker, M. (2011). Voice assimilation in French obstruents: Categorical or gradient. *Tones and features: A festschrift for Nick Clements*, 149–175.
- Halle, M., Hughes, G. W., & Radley, J. P. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29(1), 107–116.
- Haselager, G. J. T., Slis, I. H., & Rietveld, A. C. M. (1991). An alternative method of studying the development of speech rate. *Clinical Linguistics & Phonetics*, 5(1), 53–63.
- Hon, E. A. (2005). *An acoustic analysis of labialization of coronal nasal consonants in American English*. Doctoral dissertation, Massachusetts Institute of Technology.
- Hawkins, S. (2011). Does phonetic detail guide situation-specific speech recognition. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 9–18.



- Hwang, S. O. K., Monahan, P. J., & Idsardi, W. J. (2010). Underspecification and asymmetries in voicing perception. *Phonology*, 27(02), 205-224.
- Humes, L., Dubno, J. (2010). Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, et al. (Eds). *Springer Handbook of Auditory Research*, 34: The Aging Auditory System (pp. 211–258). New York: Springer.
- Hura, S. L., Lindblom, B., & Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, 35(1-2), 59-72.
- Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory and Cognition*, 12(6), 558-567.
- Janiszewski, C. & Wyer, S. (2014). Content and process priming: A review. *Journal of Consumer Psychology*, 24(1), 96-118.
- Janse, E. (2009). Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, 125(4), 2361-2373.
- Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium*, 29-54.
- Jun, J. (1995, March). Place assimilation as the result of conflicting perceptual and articulatory constraints. *West Coast Conference on Formal Linguistics*, 14, 221-238.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp 229–254). Amsterdam: John Benjamins.
- Jusczyk, P., & Luce, P. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing*, 23(1), 2-40.
- Kawahara, S., & Garvey, K. (2015). Nasal place assimilation and the perceptibility of place contrasts. *Open Linguistics*, 1(1).
- Kuzla, C., Cho, T., & Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics*, 35(3), 301-320.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-32. <https://CRAN.R-project.org/package=lmerTest>.
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245-294.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology*, 7, 637–676.

- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44-59.
- Lee, S., & Pater, J. (2008). *Phonological inference in word recognition: Evidence from Korean obstruent nasalization*. Korea University and University of Massachusetts, Amherst.
- Li, D., & Kaiser, E. (2012). Accounting for Speech Rate in Spoken Word Recognition. *Interspeech Conference*, 2009-2012.
- Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon*. Doctoral Dissertation, Indiana University.
- Magnuson, J., Dixon, J., Tanenhaus, M., & Aslin, R. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31(1), 133-156.
- Magnuson J, McMurray B, Tanenhaus M, Aslin R. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science*, 27(2), 285-98.
- Magnuson, J., & Nusbaum, H. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391.
- Magnuson, J., Tanenhaus, M., Aslin, R., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2), 202.
- Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language*, 274-284.
- Malécot, A., Johnston, R., & Kizziar, P. A. (1972). Syllabic rate and utterance length in French. *Phonetica*, 26(4), 235-251.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, 101(4), 653.
- Maye, J., Aslin, R., & Tanenhaus, M. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543-562.
- McClelland, J. & Elman, J. (1986). The TRACE model of speech perception, *Cognitive Psychology*, 18(1), 1-86.
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception?. *Trends in Cognitive Sciences*, 10(8), 363-369.
- McDonough, K., & Trofimovich, P. (2008). *Using priming methods in second language research*. Taylor & Francis.

- McMurray, B., Clayards, M., Tanenhaus, M., & Aslin, R. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064-1071.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), 33-42.
- McQueen, J. M. (2003). The ghost of Christmas future: didn't Scrooge learn to be good?: Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science*, *27*(5), 795-799.
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp.37–53). Oxford: Oxford University Press.
- McQueen, J. M., & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes*, *16*, 469–490.
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661-671.
- Meeuwissen, M., Roelofs, A., & Levelt, W. J. (2004). Naming analog clocks conceptually facilitates naming digital clocks. *Brain and Language*, *90*(1), 434-440.
- Meunier, F., & Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, *41*(3), 327-344.
- Mitterer, H. (2011). Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics*, *39*(3), 298-303.
- Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 496.
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics*, *65*(6), 956–969.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*(1), 73-103.
- Mitterer, H., Kim, S., & Cho, T. (2013). Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language*, *69*(1), 59-83.
- Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 244-263.

- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology*, *59*(8), 1395-1424.
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, *30*(3), 451-479.
- Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*(4), 527-545.
- Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language*, *59*(1), 133-152.
- Mohr, B., & Wang, W. Y. (1968). Perceptual distance and the specification of phonological features. *Phonetica*, *18*(1), 31-45.
- Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalisation. *Language and Cognitive Processes*, *25*(6), 808-839.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226.
- Nguyen, N. (2012). Representations of speech sound patterns in the speaker's brain: Insights from perception studies. *The Oxford Handbook of Laboratory Phonology*, Blackwell, Oxford, 359-368.
- Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In *Papers in laboratory phonology II: Gesture, segment, prosody*, ed. Gerard Docherty and D. Robert Ladd, 261-280. Cambridge: Cambridge University Press.
- Nolan, F., & Kerswill, P. E. (1990). The description of connected speech processes. In Ramsaran S. (Ed.). *Studies in the pronunciation of English*, 295-316. London: Routledge.
- Norris, D. (1992). Connectionism: A new breed of bottom up model. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist Approaches to Natural Language Processing*. Hove: Erlbaum.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204-238.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355-376.
- Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech*, *36*(2-3), 155-170.

- Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence* (pp. 41–60). Cambridge: Cambridge University Press.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119, 2382-2393.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3), 427-459.
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, 17(5), 525-535.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57(2), 273-298.
- Repp, B. H. (1983). Bidirectional contrast effects in the perception of VC-CV sequences. *Perception & Psychophysics*, 33(2), 147-155.
- Roelofs, A., Özdemir, R., & Levelt, W. J. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 900-913.
- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 1-33.
- Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, 20(6), 615-627.
- Scarborough, R. A. (2004). *Coarticulation and the structure of the lexicon*, Doctoral dissertation, University of California, Los Angeles.
- Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?. *Psychology and Aging*, 20(2), 261-271.
- Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas?. *Journal of Memory and Language*, 34(6), 807-820.
- Skoruppa, K., Mani, N., & Peperkamp, S. (2013). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child Development*, 84(1), 313-330.

- Smith, S. L., Pichora-Fuller, M. K., Wilson, R. H., & MacDonald, E. N. (2012). Word Recognition for Temporally and Spectrally Distorted Materials: The Effects of Age and Hearing Loss. *Ear and Hearing, 33*(3), 349-366.
- Snoeren, N. D., Gaskell, M. G., & Di Betta, A. M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 542-549.
- Snoeren, N. D., Segui, J., & Hallé, P. A. (2008). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 193-204.
- Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America, 96*(3), 1314-1324.
- Spivey, M., Joanisse, M., & McRae, K. (Eds.). (2012). *The Cambridge handbook of psycholinguistics*. Cambridge University Press.
- Steriade, D. (2001) Directional asymmetries in place assimilation: A perceptual account. In *The Role of Speech Perception in Phonology*, E. Hume & K. Johnson (eds.), New York: Academic Press, 219–250.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5<sup>th</sup> ed). Needham Height, MA: Allyn & Bacon.
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557-580.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension, *Science, 268*(5217), 1632-1634.
- Tiede, M. K., Perkell, J., Zandipour, M., & Matthies, M. (2001). Gestural timing effects in the “perfect memory” sequence observed under three rates by electromagnetometry. *The Journal of the Acoustical Society of America, 110*(5), 2657-2657.
- Tobin, S. J., Cho, P. W., Jennett, P. M., & Magnuson, J. S. (2010). Effects of anticipatory coarticulation on lexical access. *Proceedings of the Cognitive Science Society, 2200-2205*.
- Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics, 74*(6), 1284–1301.
- Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*(7-8), 979-1001.

- Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In *Proceedings of the 16th international congress of phonetic sciences*. Saarbruecken, 1949–1952.
- Vaissière, J. (2007). Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of sounds across languages. *Experimental Approaches to Phonology*, 54-71.
- Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin and Review*, 16(1), 74-79.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 1005–1015.
- Ussishkin, A., & Wedel, A. (2002). Neighborhood density and the root-affix distinction. In *Proceedings of NELS Conference*, 32(2), 539-550.
- Ussishkin, A., & Wedel, A. (2009). Lexical access, effective contrast, and patterns in the lexicon. *Perception in Phonology*, 267-292.
- Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41(3), 262-275.
- Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, 44(1), 95–118.
- Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3), 387-401.
- Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Cognition*, 128(2), 179 – 186
- Wheeldon, L. and R. Waksler. 2004. Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90(1), 401-412.
- Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology*, 40(5), 579-585.
- Winters, S. (2002). Perceptual influences on place assimilation: A case study. Invited talk, Rijksuniversiteit Groningen Tabu Day. Groningen, the Netherlands. June 28, 2002.
- Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, 38(2), 177-202.

- Wright, R. (2004). Factors of lexical competition in vowel articulation. *Papers in Laboratory Phonology VI*, 75-87.
- Zimmerer, F., Reetz, H., Lahiri, A. (2010). Place assimilation across words in running speech: corpus analysis and perception. *Journal of the Acoustical Society of America*, 125(4), 2307–2322.
- Zwitserslood, P. (1996). Form priming. *Language and Cognitive Processes*, 11(6), 589-596.



## Appendices

---

### Appendix A.

Experimental stimuli used in critical trials in Experiments 1 and 3.

<b>Prime</b>	<b>Target</b>	<b>Competitor</b>	<b>Unrelated 1</b>	<b>Unrelated 2</b>
sign	cane	came	juice	monk
swan	cone	comb	grave	raise
shine	dean	deem	walk	soft
lean	dine	dime	gulf	glove
chin	done	dumb	place	sell
bone	gain	game	tell	fill
main	grin	grim	store	fault
hen	gun	gum	tense	change
span	known	gnome	buck	end
brine	lane	lame	tick	wrong
scan	line	lime	both	hut
flan	mine	mime	doll	soon
clan	pawn	palm	sir	mesh
pain	phone	foam	act	watch
lawn	run	rum	black	seal
pin	rune	room	cake	leach
tan	sane	same	knock	word
nun	scene	seem	lose	fish
rain	screen	scream	fund	will
pun	skin	skim	front	large
coin	sun	sum	wall	gene
loon	teen	team	lay	round
bean	then	them	surf	slow
tune	win	whim	spy	pair

**Appendix B.**

Experimental stimuli used in critical trials in Experiments 2 and 4.

<b>Prime</b>	<b>Target</b>	<b>Competitor</b>	<b>Unrelated 1</b>	<b>Unrelated 2</b>
height	ate	ape	stock	dig
coat	beat	beep	fool	cab
root	cat	cap	freak	dodge
rat	cheat	cheap	hive	face
shoot	cot	cop	bell	pup
knit	cut	cup	tool	lock
neat	great	grape	rock	jet
mute	grit	grip	weed	jug
gut	hit	hip	book	vase
spit	hot	hop	dove	cash
chat	lit	lip	bed	mud
cite	loot	loop	but	jar
fit	mat	map	geek	stare
boat	pot	pop	cheek	skill
nut	seat	seep	bank	wax
clot	sheet	sheep	food	hose
wheat	shot	shop	hell	witch
dot	sit	sip	clown	gull
flat	skit	skip	truck	knife
hate	sleet	sleep	jaw	dice
kite	suit	soup	weak	boss
heat	tight	type	lax	van
fat	white	wipe	star	key
bat	wit	whip	term	deep

## Appendix C.

Experimental stimuli used in filler trials in Experiments 1-4.<sup>14</sup>

<b>Prime</b>	<b>Target</b>	<b>Competitor</b>	<b>Unrelated 1</b>	<b>Unrelated 2</b>
regime/ship	horse	sport	deaf	death
time/drop	rash	gauge	kiss	king
roam/keep	beer	tale	bash	bath
fame/soap	case	aunt	buck	buzz
slam/chirp	shame	gate	beige	bail
name/stop	bloom	bait	heave	heath
dream/trip	firm	pack	goof	goose
mom/scalp	doom	spall	lug	love
eye	aim	brat	kick	kill
toll	chrome	catch	rove	rose
try	cream	mock	with	wish
class	jam	fine	tease	teach
thick	hang	spill	barn	balm
pose	blah	spook	tin	trim
die	too	drum	warn	warm
knee	bay	has	train	tame
how	bear	gray	grain	gram
thus	mousse	antic	plan	plum
tear	coil	apex	shrine	crime
piece	brake	baton	torn	dorm
severe	greek	fast	turn	perm
close	few	matter	worn	worm
law	verse	scope	scone	scum
mystery	science	far	stain	steam
hazard	upper	habit	pill	pillar
toss	here	debit	widow	window
soda	modify	dough	east	easter
tag	after	willow	bleach	breech
dark	so	across	anchor	anger
bug	cliff	behind	wink	wind
bus	brew	like	pass	pacify
rice	way	quiet	part	apart
young	practice	first	approve	prove
leave	rather	where	explode	expose

<sup>14</sup> The first eight word pairs in the Prime list (separated by “/”) are words ending in labial nasal or oral stops that were used in the acoustic analyses reported in Chapter 2. The words ending in the labial nasal were used as fillers in Experiments 1 and 3 and the words ending in the labial oral stop were used in Experiments 2 and 4.

## Appendix C. (continued)

<b>Prime</b>	<b>Target</b>	<b>Competitor</b>	<b>Unrelated 1</b>	<b>Unrelated 2</b>
zebra	ever	light	bit	rabbit
laugh	shall	flash	phase	phrase
tack	male	back	hail	nude
ale	sound	whale	surround	park
may	rack	maple	track	double
glory	scarce	gory	scales	sofa
lacing	winter	razing	winder	rise
creaky	field	creepy	feared	tab
toy	graph	joy	grief	point
whisper	patch	whisker	badge	help
daily	pitch	daisy	pinch	ride
wearry	cigar	eerie	seagull	secret
half	ruby	calf	roomy	last
break	for	braked	fur	cry
glass	pencil	mug	desk	attest
paper	pitcher	school	chalice	allow
think	dog	brain	collar	buy
mouse	dish	cheese	kitchen	bless
leaf	prize	tree	gift	deny
trophy	clock	reward	late	token
conceal	honor	disguise	praise	mask
long	bury	high	hide	regard
dull	snail	dim	turtle	sharp
gecko	blue	lizard	color	reptile
summer	mumble	season	accent	vacation
anxiety	snow	stress	ice	utter
small	thigh	sight	might	night
sole	through	taught	thought	though
indeed	number	lumber	hamster	dumber
tooth	frizzle	frizzy	frozen	thrown
mode	subtle	sutton	sublet	include
crazy	wedging	wedding	when	float
averse	choose	adverse	adverb	welcome
life	branch	file	still	split
circular	tall	circle	bicycle	cycle
colonel	beauty	corner	colon	port
work	do	woke	worse	wore
sense	outer	cent	scents	appeal

**Appendix D.**

Word frequency retrieved from SUBTLEXus database for the displayed prime, undisplayed prime competitor (if available), and displayed target and displayed target competitor words used in Experiments 1-4.

<b>Prime</b>	<b>Freq.</b>	<b>Prime Comp</b>	<b>Freq</b>	<b>Target</b>	<b>Freq</b>	<b>Target Comp.</b>	<b>Freq.</b>
bean	6.84	beam	8.73	scene	74.65	seem	139.82
clan	4.1	clam	3.92	pawn	4.33	palm	13.24
hen	3.2	hem	0.82	gun	213.2	gum	13.39
loon	1.41	loom	0.41	teen	4.1	team	147.61
nun	6.96	numb	4.88	then	1489.53	them	1778.82
scan	9.41	scam	6.14	line	206.18	lime	3.29
span	1.98	spam	1.16	known	123.53	gnome	0.63
tune	15.61	tomb	5.63	win	134.65	whim	2.33
bone	26.06	NA	NA	gain	13.73	game	233.84
brine	0.31	NA	NA	lane	33.41	lame	10.92
chin	12.69	NA	NA	done	485.04	dumb	46.96
coin	9.75	NA	NA	sun	69.67	sum	10.25
flan	1.69	NA	NA	mine	250.98	mime	1.14
lawn	12.35	NA	NA	run	350.55	rum	7.49
lean	10.37	NA	NA	dine	4.29	dime	12.06
main	42.73	NA	NA	grin	2.71	grim	4.78
pain	97.94	NA	NA	phone	269.73	foam	3.51
pin	16.37	NA	NA	rune	0.98	room	439.51
pun	1.84	NA	NA	skin	44.04	skim	0.84
rain	48.9	NA	NA	screen	23.39	scream	26.41
shine	20.65	NA	NA	dean	38.71	deem	0.59
sign	133.27	NA	NA	cane	8.33	came	463.73
swan	6.82	NA	NA	cone	2.92	comb	6.06
tan	8.61	NA	NA	sane	5.51	same	417.18

**Appendix D.** *(continued)*

<b>Prime</b>	<b>Freq.</b>	<b>Prime Comp</b>	<b>Freq</b>	<b>Target</b>	<b>Freq</b>	<b>Target Comp.</b>	<b>Freq.</b>
chat	16.27	chap	6.35	lit	9.73	lip	10.75
coat	42.08	cope	3.25	beat	131.69	beep	6.51
flat	26.22	flap	2.65	skit	0.82	skip	21.1
heat	40.08	heap	4.88	tight	50.92	type	60.65
height	9.33	hype	1.63	ate	33.76	ape	9.67
knit	1.9	nip	3.24	cut	229.76	cup	51.65
rat	32.61	rap	13.04	cheat	17.67	cheap	36.24
wheat	5.75	weep	5.49	shot	227.43	shop	53.55
hate	214.59	NA	NA	sleet	0.65	sleep	227.94
mute	2.49	NA	NA	grit	0.94	grip	9.69
gut	9.75	NA	NA	hit	275	hip	15.41
root	10.47	NA	NA	cat	66.33	cap	18.75
shoot	164.94	NA	NA	cot	2	cop	86.14
clot	1.8	NA	NA	sheet	11.61	sheep	13.43
dot	6.63	NA	NA	sit	311.35	sip	5.1
kite	2.29	NA	NA	suit	68.61	soup	25.2
boat	95.78	NA	NA	pot	22.53	pop	67.47
cite	0.94	NA	NA	loot	3.55	loop	6.76
neat	12.45	NA	NA	great	820.86	grape	4
fat	79.43	NA	NA	white	171.45	wipe	16.88
fit	57.59	NA	NA	mat	3.49	map	31.82
bat	20.63	NA	NA	wit	5.92	whip	13.16
spit	19.35	NA	NA	hot	189.84	hop	19.16
nut	15.63	NA	NA	seat	78.78	seep	0.18