Individual differences in plasticity in speech perception

Donghyun Kim

Department of Linguistics

McGill University

July 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

The current dissertation addresses individual differences in plasticity in speech perception under adverse conditions. Although speech sounds are highly variable and listeners may have limited cognitive resources under challenging conditions, listeners are shown to overcome these challenges by flexibly adapting to varying contextual demands and achieve perceptual constancy. As such, understanding how listeners cope with adverse conditions is a fundamental issue for better understanding the perceptual processes and cognitive mechanisms underlying speech perception. Most studies of perceptual adaptation in speech have thus far typically reported on group level observations, which may mask considerable differences in how individual listeners adapt. This dissertation addresses these issues by examining the mechanisms underlying adaptive plasticity in speech perception using an individual differences approach.

Study 1 examines whether listeners flexibly adapt to unfamiliar speech patterns such as those encountered in foreign-accented English vowels. In these cases, the relative informativeness of acoustic dimensions (spectral quality vs. duration) can be changed such that the most informative dimension (spectral quality) is no longer informative, but the role of the secondary dimension (duration) is enhanced. Results showed that listeners flexibly adapt to unfamiliar speech sounds by increasing reliance on less informative acoustic dimensions when the most informative acoustic dimension is no longer useful. Results also showed that individual listeners varied widely in patterns of perceptual adaptation and these differences in adaptive strategies were linked to individual differences in cognitive abilities (i.e., inhibitory control).

Study 2 investigates how and to what extent speech perception abilities are modulated under cognitive load and whether individuals differ in the extent to which they adjust their cue weighting strategies in the utilization of multiple acoustic cues to cope with this adverse condition. Results revealed that listeners overall showed increased cue weights, which may be interpreted as compensatory cue weighting strategies to adapt to phonetic categories under cognitive load. However, there were large individual differences in the extent to which these compensatory strategies manifested and these differences were associated with individual listeners' cognitive abilities (i.e., working memory and inhibitory control).

Taken together, the present dissertation showed that listeners are remarkably flexible in adapting their phonetic categories to cope with challenging conditions. This dissertation also showed that individual listeners differed substantially in the extent to which they made adjustments of their cue weighting strategies under adverse conditions and these differences are related to higher-level cognitive functions. These findings provide insights into the interplay between speech perception and cognitive processes, leading to a more comprehensive understanding of the mechanisms underlying plasticity in speech perception.

Résumé

La thèse actuelle aborde les différences individuelles en ce qui concerne la malléabilité dans la perception de la parole dans des conditions défavorables. Bien que les sons de la parole sont très variables et les auditeurs peuvent avoir des ressources cognitives limitées dans des conditions difficiles, les auditeurs démontre de l'aptitude pour surmonter ces défis en s'adaptant avec souplesse aux différentes exigences contextuelles et parvennent à avoir une perception cohérente. En tant que tel, comment les auditeurs font face à des conditions défavorables est une question fondamentale pour obtenir une meilleure compréhension des processus perceptifs et des mécanismes cognitifs sous-jacents à la perception de la parole. La plupart des études sur la malléabilité de la perception de la parole ont jusqu'à présent fait l'objet d'observations de groupe, ce qui peut masquer des différences considérables dans la façon dont les auditeurs individuels s'adaptent. Cette dissertation aborde ces problèmes en examinant les mécanismes sous-jacents de la plasticité adaptative dans la perception de la parole en utilisant une approche se basant sur des différences individuelles.

L'étude 1 examine si les auditeurs s'adaptent avec souplesse à des modèles de la parole non familiers tels que ceux rencontrés dans les voyelles anglaises ayant un accent étranger. Dans ces cas, la valeur informative relative des dimensions acoustiques (qualité spectrale par rapport à la durée) peut être modifiée de telle sorte que la dimension la plus informative (qualité spectrale) n'est plus informative, mais le rôle de la dimension secondaire (durée) est optimisé. Les résultats ont montré que les auditeurs s'adaptaient avec souplesse aux sons de la parole qui ne leur sont pas familiers en se basant davantage sur des dimensions acoustiques moins informatives lorsque la dimension acoustique la plus informative n'est plus utile. Les résultats ont également montré que les auditeurs individuels variaient largement dans les modèles d'adaptation perceptive et que ces différences dans les stratégies adaptatives étaient liées aux différences individuelles dans les capacités cognitives (c.-à-d., le contrôle inhibiteur).

Étude 2 examine comment et dans quelle mesure les capacités de perception de la parole sont sous la charge cognitive et modulée et si les individus se diffèrent dans la mesure à laquelle ils ajustent leurs stratégies de pondération des repères perceptifs lors de l'utilisation de multiples signaux acoustiques pour faire face à cette situation défavorable. Les résultats ont révélé globalement que des auditeurs ont augmenté la valeur des repères perceptifs, ce qui peut être interprété comme une évidence de l'utilisation des stratégies de pondération des signaux compensatoires afin de s'adapter aux catégories phonétiques sous la charge cognitive. Cependant, il y avait de grandes différences individuelles dans la mesure dans laquelle ces stratégies compensatoires se sont manifestées et ces différences individuelles étaient associées avec les capacités cognitives des auditeurs (c.-à-d., la mémoire de travail et de contrôle inhibiteur).

En somme, la présente dissertation a démontré que les auditeurs possèdent une malléabilité remarquable dans l'adaptation de leurs catégories phonétiques afin de faire face à des conditions difficiles. Cette dissertation a également indiqué que les auditeurs individuels se diffèrent substantiellement dans la mesure à laquelle ils ont fait des ajustements de leurs stratégies de pondération de repères dans des conditions défavorables et ces différences sont liées à des fonctions cognitives supérieures. Ces résultats donnent un aperçu de l'interaction entre la perception de la parole et les processus cognitifs, menant à une compréhension plus complète des mécanismes sous-jacents de la malléabilité de la perception de la parole.

Acknowledgements

Writing this dissertation would not have been possible without the help and support from many people. First and foremost, I would like to thank my advisor, Dr. Meghan Clayards, for her continuous support and encouragement throughout my graduate school years. She provided me with invaluable academic and personal guidance during my PhD studies and gave me the inspiration which eventually led to the topic of this dissertation.

I would also like to thank the other members of my dissertation committee, Drs. Morgan Sonderegger and Eun Jong Kong, who gave me vital comments and insightful suggestions for my thesis. I further wish to thank Dr. Heather Goad for her academic guidance and mentoring. I especially learned a great deal of academic writing from her comments and feedback on my first evaluation paper at McGill.

I would also like to thank my mentors in South Korea, Drs. Minpyo Hong and Jae-Young Lee, for their unwavering support and encouragement over the years. Their courses in linguistics initially sparked my interest in linguistics and their encouragement helped me go through my graduate school years.

Last but not least, I would like to thank my family and friends for their constant support, love, encouragement and friendship. I would never have completed this project without them.

Contribution of Authors

Chapter 1: General introduction
Donghyun Kim solely authored this chapter, with feedback and comments from Meghan Clayards and Morgan Sonderegger.

Chapter 2: Study 1
This manuscript has been prepared for journal submission. It is co-authored by Meghan Clayards and Eun Jong Kong. Donghyun Kim performed the literature review, designed the experiments, collected the data, analyzed the data, and drafted the manuscript. Meghan Clayards and Eun Jong Kong provided guidance during the development of the experiments and provided comments on drafts of the manuscript. Morgan Sonderegger provided comments on drafts of the manuscript.

Chapter 3: Study 2
This manuscript has been prepared for journal submission. It is co-authored by Meghan Clayards. Donghyun Kim performed the literature review, designed the experiments, collected the data, analyzed the data, and drafted the manuscript. Meghan Clayards provided guidance during the development of the experiments and provided comments on drafts of the manuscript. Morgan Sonderegger provided comments on drafts of the manuscript.

Chapter 4: General discussion and conclusion
Donghyun Kim solely authored this chapter, with feedback and comments from Meghan Clayards and Morgan Sonderegger.

Contents

Abstract **2**

Résumé **3**

Acknowledgments **4**

Contribution of Authors **5**

Chapter 1 **10**
General introduction **10**
1.1 Speech perception in adverse conditions ... 10
1.2 Flexibility in speech perception in adverse conditions ... 12
1.3 Individual differences in speech perception ... 14
1.3.1 Sources of individual differences in speech perception ... 15
1.4 Overview of the dissertation ... 18

Chapter 2 **19**
Study 1: Individual differences in perceptual adaptation to phonetic categories **19**
2.1 Introduction ... 19
2.1.1 Flexibility in speech perception ... 19
2.1.2 Individual differences in perception of acoustic cues to speech ... 23
2.1.3 Cognitive abilities in speech perception ... 24
2.1.4 Categorization gradiency in speech perception ... 27
2.1.5 The present study ... 28
2.2 Methods ... 30
2.2.1 Participants ... 30
2.2.2 Stimuli ... 30
2.2.3 Procedure ... 32
2.2.4 Analysis ... 34
2.3 Results ... 36
2.3.1 Vowel categorization at Baseline 1 ... 36
2.3.2 VAS and cognitive measures ... 37
2.3.3 The adaptation task ... 42
2.4 Discussion ... 46
2.4.1 Perceptual adaptation to unfamiliar speech ... 47
2.4.2 Relationship between gradiency and secondary cue use ... 49
2.4.3 The role of categorization gradiency in perceptual adaptation ... 49
2.4.4 The role of cognitive abilities in perceptual adaptation ... 50
2.4.5 Working memory for speech and adaptation ... 52
2.5 Conclusion ... 52

Preface to Chapter 3 **54**

Chapter 3 55
Study 2: Individual differences in flexibility in speech perception under cognitive load 55
3.1 Introduction 55
3.1.1 Speech perception in adverse conditions 56
3.1.2 The effect of cognitive load on speech perception 56
3.1.3 The role of cognitive abilities in speech perception 59
3.1.4 The role of gradiency in speech perception 62
3.1.5 The present study 63
3.2 Methods 64
3.2.1 Participants 64
3.2.2 Stimuli 65
3.2.3 Procedure 67
3.2.4 Analysis 69
3.3 Results 72
3.3.1 Individual difference measures 72
3.3.2 The effect of cognitive load on speech perception in relation to individual difference measures 75
3.4 Discussion 82
3.4.1 The effect of cognitive load on speech perception 82
3.4.2 The role of cognitive abilities in speech perception under cognitive load 85
3.4.3 The role of categorization gradiency in speech perception under cognitive load 87
3.5 Conclusion 88

Chapter 4 89
General discussion and conclusion 89
4.1 Summary of the dissertation 89
4.2 Implications for speech perception 91
4.3 Future directions 93
4.4 Conclusion 95

References 96

List of Tables

Table 2.1. Correlation matrix between cognitive ability measures ……………………….…... 40

Table 2.2. Summary of linear regression model predicting categorization gradiency. Model coefficient estimates (β), standard errors (SE), corresponding *t*-values, and *p*-values …….. 41

Table 2.3. Summary of perceptual adaptation model. Model coefficient estimates (β), standard errors (SE), corresponding *z*-values, and *p*-values ……………………………….. 44

Table 3.1. Correlation matrix between cognitive ability measures ……………………….... 73

Table 3.2. Summary of cognitive load model. Model coefficient estimates (β), standard errors (SE), corresponding *z*-values, and *p*-values ……………………………………….……. 80

List of Figures

Figure 2.1. Illustration of the stimuli used in the VAS and the adaptation task. Test stimuli for the adaptation task were the red square and the blue triangle 31

Figure 2.2. Hypothetical illustration of duration cue use at Baseline 1 as measured by the difference in 2AFC crossover points between short and long vowel durations 35

Figure 2.3. Proportion of /æ/ responses along vowel spectral quality continuum at Baseline 1 as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are logistic curves fit to each individual listener data for each vowel duration 37

Figure 2.4. A histogram of overall visual analog scaling (VAS) responses 38

Figure 2.5. Visual analog scaling (VAS) responses for three representative listeners (more categorical vs. more gradient). The VAS slope values (gradiency measure) of each representative listener are 132 (Participant 187), 55 (Participant 192), and 10 (Participant 188) in which smaller values indicate more gradient responses 39

Figure 2.6. Categorization gradiency as a function of secondary cue use and working memory. The shallower slopes (smaller values) indicate more gradient responses. Higher Corsi scores indicate better working memory capacity 42

Figure 2.7. (A) Proportion /æ/ vowel responses of the Test stimuli across blocks as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are individual listeners' Test stimulus responses across blocks. (B) The effect of individual differences in Stroop interference (inhibitory control) on categorization responses for short (130 ms) and long vowel (330 ms) durations across blocks. A higher Stroop interference score indicates low inhibitory control, and each number indicates individual listeners 45

Figure 3.1. Illustration of the stimuli in the VAS and the auditory stimuli in the dual task ... 66

Figure 3.2. Example stimuli for the concurrent visual search task. The left panel contains an oddball object (black diamond) and the right panel does not 66

Figure 3.3. Visual analog scaling (VAS) responses by two sample listeners, namely Listener 255 (more categorical) and Listener 259 (more gradient) 75

Figure 3.4. (A) Heatmap plots of overall responses for spectral quality and duration under no load and cognitive load conditions. (B) Proportion of /æ/ responses as a function of vowel spectral quality and duration differences for the no load and cognitive load conditions. Thin lines are logistic curves fit to each individual listener's data 77

Figure 3.5. Proportion of /æ/ responses along vowel spectral quality continuum as a function of vowel duration differences by three listeners 79

Figure 3.6. Scatter plots of individual listeners' spectral cue weight differences between cognitive load and no load conditions as a function of Go/No-go scores (A) and duration cue weight differences between cognitive load and no load conditions as a function of working memory composite scores (B) 81

Chapter 1

General introduction

1.1 Speech perception in adverse conditions

Speech communication does not always occur in ideal listening conditions. It is in fact rarely the case that listeners hear speech in quiet and fully focused conditions as in the laboratory setting. In everyday life, it is common that listeners recognize speech from noisy environments or unfamiliar speech that deviates from stored speech patterns due to foreign accents, dialects, or speech disorders. A variety of factors contribute to these adverse conditions that listeners face in real-life speech communication. According to Mattys, Davis, Bradlow, and Scott (2012), these factors are largely categorized into talker-related challenges (e.g., accented speech, non-native accents, or disordered speech), environmental degradations (e.g., background noise or competing speech), and listeners' limited processing resources (e.g., cognitive load, dual-tasking, or hearing impairments). Although listeners initially have reduced comprehension of speech under these adverse conditions, they have been shown to overcome these challenges by flexibly accommodating and adapting to the demands of these conditions (e.g., Guediche, Blumstein, Fiez, & Holt, 2014; Mattys et al., 2012).

Given that speech communication in adverse conditions is an essential part of everyday life, understanding how listeners cope with these conditions is a fundamental issue for better understanding the mechanisms underlying adaptive plasticity in speech perception. A considerable body of research has investigated speech perception in different challenging conditions (Bradlow & Bent, 2008; Cainer, James, & Rajan, 2008; Clarke & Garrett, 2004; Kraljic & Samuel, 2005; Mattys & Palmer, 2015; Mattys & Wiget, 2011; Norris, McQueen,

& Cutler, 2003). For example, previous work has shown that only brief exposure to an accented speaker allows listeners to adapt to the speaker's accent, leading to intelligibility improvements for the accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004). Such improvements in speech intelligibility were also observed when listeners accommodate speech in background noise (Cainer et al., 2008). In addition to intelligibility improvements, a number of studies have shown that listeners adapt to atypical pronunciations by using stored linguistic (e.g., lexical) knowledge (Kraljic & Samuel, 2005; Norris et al., 2003). Previous research has also found that listeners have limited attentional resources and that increased cognitive load under dual-task conditions has a modulating effect on speech processing (Mattys & Palmer, 2015; Mattys & Wiget, 2011).

Under ideal listening conditions, participants' performance may reach ceiling or substantial individual differences may be less likely to be uncovered. However, under adverse conditions, listeners may reveal differential compensatory strategies to cope with challenges in recognizing speech. These potential differences across individuals may not be random and likely reflect important differences in the mechanisms underlying speech perception. In this regard, adverse conditions are particularly useful for investigating individual differences in speech perception because meaningful differences across individuals (e.g., compensatory strategies) are likely to emerge under sub-optimal conditions. Also, because the findings in previous studies have been mostly based on group-level observations, relatively little has been understood about how listener-specific characteristics such as cognitive abilities and within-category sensitivity in phoneme categorization are related to individual differences in speech perception in adverse conditions. Furthermore, although much attention has been paid to perceptual adaptation in speech based on top-down lexical knowledge, relatively little has been reported on adaptability of speech sound categories based on bottom-up acoustic-phonetic information.

The overarching goal of the current dissertation is to better understand the mechanisms underlying adaptive plasticity in speech perception. To that end, this dissertation investigates whether listeners adapt their use of acoustic cues in phonetic categorization in two different challenging conditions—source degradation and listener limitations in adaptability of phonetic categories—both of which commonly occur in everyday listening conditions but have relatively received less attention than environmental degradations (e.g., speech in background noise). The present dissertation investigates how individuals adapt to these challenging listening conditions in two studies. Study 1 aims to simulate a common source degradation condition (i.e., foreign-accented speech) and examines listeners' adaptability of foreign-accent-like sound categories. Study 2 addresses an adverse condition arising from receiver limitations (i.e., cognitive load) and investigates how listeners cope with cognitive load in speech sound categorization.

1.2 Flexibility in speech perception in adverse conditions

Despite the challenges listeners may encounter in everyday listening conditions, they flexibly adapt to these challenging conditions without much difficulty. Much interest has recently been paid to this flexibility of speech perception and perceptual learning in speech (Idemaru & Holt, 2011; Kraljic & Samuel, 2005; Norris et al., 2003, among many others). One mechanism that has been suggested in these studies is a perceptual learning mechanism based on top-down lexical knowledge (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Norris et al., 2003). These studies indicated that listeners rapidly adapt to unfamiliar or ambiguous speech input using existing lexical knowledge. This adaptation is accomplished by adjusting phonetic category boundaries in response to variation in the speech input. For example, Norris et al. (2003) showed that when listeners encounter a talker whose acoustic realization of /f/ is ambiguous between [f] and [s], listeners make a short-term adjustment to their category boundary to perceive the ambiguous stimulus as /s/ (e.g., hearing cactu[f] for cactus, namely an /s/-final word with no /f/-final counterpart). This flexible retuning of phonetic category boundaries in speech perception is termed *lexically-guided perceptual learning* (Norris et al., 2003) and helps listeners cope with talker-related challenging conditions by rapidly adapting to patterns of variation in the incoming speech. This process of perceptual learning indicates that listeners are sensitive to acoustic-phonetic information in the input signal in a given lexical context and they make short-term adjustments of phonetic categories to resolve ambiguity. These studies indicated that listeners compensate for adverse listening conditions utilizing top-down lexical information.

More relevant to the current dissertation is flexibility in speech perception based on reorganization of patterns of acoustic-phonetic cues to phonological categories (Francis, Kaganovich, & Driscoll-Huber, 2008; Idemaru & Holt, 2011; 2014). A phonological contrast has multiple acoustic-phonetic cues and listeners use these multiple cues when they perceive a speech sound contrast. Studies have shown that each phonological category is signaled by multiple acoustic-phonetic cues, and listeners give more attention to some acoustic cues over others in phonological categorization (Francis et al., 2008; Holt & Lotto, 2006). Although listeners are generally sensitive to the distribution of acoustic cues defining phonological categories, the distributional information of acoustic cues in the speech signal varies substantially due to different realizations of speech sounds such as accents and dialects.

Recent studies have documented how listeners' adaptive utilization of multiple acoustic-phonetic cues contributes to adaptability of sound categories (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015; Schertz, Cho, Lotto, & Warner, 2016). This suggests that adaptation of speech sounds can be enabled by the use of bottom-up acoustic processing analyzing distributional properties of the input signal. In particular, Holt and colleagues (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015) have suggested that phonetic category restructuring can occur based on bottom-up category internal information, which was referred to as *dimension-based statistical learning* in their work. According to this paradigm, listeners reorganize acoustic-phonetic structure of non-canonical exemplars in the speech signal, which deviates from the listeners' long-term representations of those sounds. They created an artificial accent by reversing the correlation between two acoustic dimensions from long-term native language regularities. They found that listeners rapidly adapt to the accented speech. That is, listeners down-weighted reliance on a secondary acoustic dimension (e.g., vowel duration) for vowel categorization when the correlation between the primary dimension (e.g., spectral quality) and the secondary dimension was reversed relative to the long-term English norm.

It is plausible that dimension-based statistical learning plays a role in flexibility in speech perception. In other words, the adjustment of cue weighting strategies to adapt to contextual demands would contribute to speech perception in difficult listening conditions. This adaptability of speech sound categories is particularly relevant to the current thesis in that listeners are expected to modify their cue weighting strategies when they encounter atypical exemplars in the input and also under increased cognitive demands to adapt to varying contextual demands (Francis et al., 2008). Based on bottom-up acoustic-phonetic processing, the current dissertation addresses plasticity in speech perception in terms of changes in cue weighting strategies to compensate for adverse conditions. Study 1 will examine whether listeners flexibly adapt to unfamiliar speech sounds that deviate from native language patterns by making short-term changes to acoustic cues, and Study 2 will investigate whether listeners adjust their cue weighting strategies under cognitive load.

1.3 Individual differences in speech perception

Although a great deal of research on speech perception has reported on group-level observations, recent studies have indicated that individuals differ in how they perform speech perception tasks (e.g., Hazan & Rosen, 1991). It has also been documented that these individual differences are stable across testing sessions (Idemaru, Holt, & Seltman, 2012; Kong & Edwards, 2016; Schertz, Cho, Lotto, & Warner, 2015; Yu & Lee, 2014), suggesting that they largely reflect stable listener attributes rather than random variability. These previous studies have mostly investigated in optimal listening conditions. However, it is likely that larger individual differences in speech perception will emerge under sub-optimal conditions. This is because in optimal conditions a number of individuals tend to perform at ceiling and it is possible that different compensatory strategies are adopted to cope with challenging conditions (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009; Mattys et al., 2012).

Understanding the nature of and factors contributing to such individual differences is crucial for better understanding the mechanisms underlying speech perception processes and their potential interplay with other listener characteristics such as individual cognitive abilities. In particular, speech perception in challenging conditions is an ideal ground for testing individual differences in speech perception. Recent studies have provided some evidence of potential sources of individual differences in speech perception. One possible source is differences in cognitive abilities underlying speech perception processes. It has been suggested that general cognitive abilities such as working memory, attention, and inhibition aid more general learning processes (Goldstone, 1998). Previous research has suggested that individual differences in general cognitive abilities play a role in the perception of speech sounds (Akeroyd, 2008). In the current thesis, two potential sources of individual differences in speech perception will be considered, i.e., domain-general cognitive abilities and gradiency in phoneme categorization, which will be described in the following sections.

1.3.1 Sources of individual differences in speech perception

1.3.1.1 Cognitive abilities

An emerging body of research has indicated that listeners utilize cognitive resources for speech perception especially in challenging listening conditions (e.g., Akeroyd, 2008;

Arlinger et al., 2009; Rönnberg et al., 2013). These studies have found that speech perception performance under challenging conditions can be explained to some extent by individual differences in cognitive abilities. These cognitive abilities have been commonly measured by key components of executive functions including working memory, inhibition, and cognitive flexibility (Diamond, 2013; Friedman & Miyake, 2017; Miyake & Friedman, 2012) and have been shown to account for some of the variance in speech perception. Previous work has used these executive function measures to test the effect of cognitive abilities on speech perception processes (Banks, Gowen, Munro, & Adank, 2015; Janse, 2012; Kim & Hazan, 2010; Kong & Edwards, 2016; Rönnberg et al., 2013; Souza & Arehart, 2015). For example, individuals with better working memory capacity have been observed to show better speech perception performance in degraded listening conditions (Rönnberg et al., 2013; Souza & Arehart, 2015). Inhibition has also been observed to play a role in perceptual adaptation to unfamiliar accents (Banks et al., 2015) and competing speech (Janse, 2012).

However, findings of possible links between individual differences in speech perception and cognitive flexibility have been mixed (Kim & Hazan, 2010; Kong & Edwards, 2016). For example, Adank and Janse (2010) found that cognitive flexibility predicted differences in comprehension of a novel accent in older adults, but Kim and Hazan (2010) did not find a link between perceptual learning of novel sounds and cognitive flexibility. In addition to the core executive functions, this thesis also probes a possible link between perception of speech and sustained attention, which assesses individuals' ability to maintain attention for a certain amount of time (Jongman, Roelofs, & Meyer, 2015). This attentional measure was included to test whether the ability to maintain attention during learning tasks would predict the patterns of perceptual learning in speech.

1.3.1.2 Phoneme categorization gradiency

Another potential source of individual differences in speech perception is gradiency in phoneme categorization. Studies in recent years have shown that individuals differ in how gradiently they perceive speech sounds such that some listeners show gradient patterns of phoneme categorization while others show categorical patterns (Kapnoula, Winn, Kong, Edwards, & McMurray, 2017; Kong & Edwards, 2011; 2016; Schellinger, Munson, & Edwards, 2017). These studies have suggested that gradient categorization patterns may be useful because they allow for flexibility in how acoustic cues are mapped onto sound categories. For example, recent studies examined whether variability in categorization gradiency is related to listeners' sensitivity to multiple acoustic cues (e.g., VOT (primary) and f0 (secondary)) to the stop voicing contrast in English (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). They found that individuals who had a more gradient pattern of speech sound categorization were more sensitive to the secondary acoustic cue (i.e., f0). These findings indicate that listeners' sensitivity to gradient within-category differences may be a potential factor contributing to individual differences in adaptability of speech sound categories.

As a measure of gradiency of phoneme categorization, the aforementioned studies employed a visual analog scaling (VAS) task, a continuous measure of phonetic categorization that allows for gradient responses (Massaro & Cohen, 1983). Recent studies that used the VAS task showed that there are considerable individual differences in how much listeners show gradiency when they categorize speech sounds (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016), in which some listeners exhibited a more categorical pattern in favor of endpoint responses while others showed a more gradient pattern using a wide range of available responses. Further work showed that more gradient listeners were better able to predict upcoming words based on fine phonetic detail, suggesting they are able to maintain this information (Kapnoula, Edwards & McMurray, 2015). This ability may also be useful for adapting to adverse listening condition, for example by allowing the listener to better notice changes in patterns of cue use or by providing a richer representation to begin with. This gradiency measure of phoneme categorization will be used in both Study 1 and Study 2 to determine whether sensitivity to within-category differences is linked to adaptability of phonetic categories in perceptual adaptation to unfamiliar speech (Study 1) and speech perception under cognitive load (Study 2). Thus, the potential usefulness of maintaining fine phonetic details in speech perception will be in part tested in two different adverse conditions.

1.4 Overview of the dissertation

The rest of the dissertation is organized as follows. Chapter 2 presents Study 1 on individual differences in perceptual adaptation to phonetic categories. Chapter 3 presents Study 2 on individual differences in flexibility in speech perception under cognitive load. Chapter 4 summarizes and discusses the findings of Study 1 and Study 2, and it also concludes with implications for existing theories of speech perception and possible avenues for future research.

Chapter 2

Study 1: Individual differences in perceptual adaptation to phonetic categories

2.1 Introduction

When perceiving speech, listeners face an enormous amount of variability in speech sound realization. This variability can come from diverse sources such as degraded speech, disordered speech, or even idiosyncratic pronunciations. Also, as people travel more than ever within and across countries, it is not uncommon to converse with people who have regional dialects or foreign accents, each of which sounds unfamiliar. Although these highly variable speech sounds can be a challenge to understanding speech, it has been observed that listeners are flexible in speech recognition and rapidly adapt to unfamiliar pronunciations (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004, among many others). The goal of the present study is to better understand this remarkably flexible process in speech. In this study we present how listeners flexibly adapt to unfamiliar speech patterns such as those encountered in foreign-accented English vowels and what makes some listeners better adapters to these unfamiliar speech patterns.

2.1.1 Flexibility in speech perception

Even pronunciations of one speech sound of a language can vary widely depending on dialects, accents, gender differences, idiosyncratic differences, and even from instance to instance (e.g., Newman, Clouse & Burnham, 2001). Despite this variability in speech sound realization, listeners can often overcome initial difficulties and show intelligibility improvements with relatively brief exposure to this highly variable input (e.g., Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004). For example, listeners show improvements in category identification accuracy (Baese-Berk et al., 2013; Bradlow & Bent, 2008) and in processing speed (Clarke & Garrett, 2004) after they become familiar with foreign accented speech. These studies have shown that this adaptation process to unfamiliar accents occurs quite rapidly with relatively brief exposure. A considerable body of literature has examined this flexibility in speech perception in terms of how perceptual systems are able to adapt rapidly and make relevant adjustment to accommodate patterns of variation in speech input (e.g., Idemaru & Holt, 2011; 2014; McQueen, Cutler, & Norris, 2006; Norris, McQueen, & Cutler, 2003, among many others). These studies have mostly focused on phonetic categories and how they are retuned to cope with acoustic-phonetic variability. Given the variability inherent in speech perception processes, understanding how listeners successfully adapt and understand speakers whose productions differ from familiar phonological patterns is an important goal in speech perception.

One set of studies has provided evidence that listeners adapt to the acoustic-phonetic variability using top-down linguistic information (Kraljic & Samuel, 2005; McQueen et al., 2006; Norris et al., 2003, among many others). These studies have demonstrated that listeners flexibly adjust phonetic category boundaries in response to variation in the speech input. For example, when listeners encounter a talker whose acoustic realization of /f/ (as in *giraffe*) is ambiguous between [f] and [s], listeners make a short-term adjustment to their category boundary to perceive the ambiguous stimulus as /f/ (Norris et al., 2003). This phonetic adjustment seems to be driven by the disambiguating lexical context (e.g., hearing gira[s/f] for giraffe, namely an /f/-final word with no /s/-final counterpart). This lexically-guided perceptual learning in speech can help listeners cope with acoustic-phonetic variability by responding to patterns of variation in the speech input. In addition, research has shown that listeners use top-down contextual information to adapt to speech variability (e.g., Bradlow & Alexander, 2007; Pichora-Fuller, 2008). For example, Pichora-Fuller (2008) showed that listeners utilize semantic context to facilitate perception of speech when there is a mismatch between speech signal and meaning.

In addition to the use of top-down linguistic knowledge, perceptual adaptation can also be enabled by the use of bottom-up analyses of distributional properties of the input speech signal (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015; Schertz, Cho, Lotto, & Warner, 2016). In particular, Idemaru and Holt (2011, 2014) have shown that phonetic category restructuring can occur based on category internal information, which they termed *dimension-based statistical learning*. In this paradigm, listeners adjust their use of the various acoustic dimensions that define speech sound categories. Idemaru and Holt (2011, 2014) used spoken words such as *pier* and *beer*, in which the initial segment varied both in voice onset time (VOT) and in pitch at vowel onset (f0). The English stop voicing contrast (e.g., /p/ vs. /b/) is primarily distinguished based on VOT, with f0 being secondary. Productions of voiceless stops generally have longer VOTs than voiced stops and voiceless stops also tend to have higher f0 than voiced stops. At baseline, VOT and f0 were correlated as they are naturally for English—high f0 associated with long VOT (the Canonical block). In the following block, the correlation between VOT and f0 was reversed—low f0 was associated with long VOT (the Reverse block). Both blocks included test stimuli that were ambiguous in VOT but were either high or low in f0 and responses to these test stimuli were compared across blocks. In the Canonical block, listeners responded /p/ much more for the high f0 than for the low f0 test stimulus indicating that this cue was being used to distinguish the contrast for these listeners. In the Reverse block, on the other hand, listeners gave equivalent responses to the high and low f0 test stimuli. This indicates that exposure to the change in the correlation of f0 with VOT led listeners to down-weight f0 in English stop voicing categorization. That is, listeners ceased to use f0. These findings suggest that listeners are well aware of the distributional properties of the speech signal involving secondary acoustic dimensions as well as primary acoustic dimensions.

Further work has extended this paradigm to other contrasts (Lehet & Holt, 2017; Liu & Holt, 2015; Schertz et al., 2016). For example, Liu and Holt (2015) examined the dimension-based statistical learning of vowels and found that at baseline native English listeners rely primarily on spectral quality with vowel duration being secondary, consistent with previous work (Hillenbrand, Clark, & Houde, 2000; Kondaurova & Francis, 2008; 2010). When exposed to an artificial accent which deviates from English norms, however, listeners flexibly down-weighted their use of vowel duration. These studies have shown that listeners use a more reliable dimension (VOT or spectral quality) as the basis for perceptual learning about the distribution of a less reliable dimension (f0 or vowel duration) in the category. The present study examines whether listeners can also use distributions to learn which dimension is the most reliable. We expect that listeners will increase their use of a secondary cue to adapt to unfamiliar pronunciations when the most informative dimension (spectral quality) is no longer informative, but the informativeness of the secondary cue (duration) is enhanced in an English vowel contrast.

The dimension-based perceptual learning paradigm is particularly relevant to the present study in that listeners are sensitive to patterns of variation in multiple acoustic cues in the speech signal. Accordingly, the perceptual adaptation task in the current study is based on this perceptual learning paradigm (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015; Schertz et al., 2016). It is expected that listeners are able to learn from and adapt to changes in the distributional properties of the input signal by adjusting their expectations about acoustic-phonetic information in that speech exemplar accordingly.

In fact, atypical speech that deviates from native language norms and requires enhancement by non-primary acoustic dimensions is not uncommon. For instance, non-native pronunciations of English front vowel contrasts (e.g., from Spanish, Korean, Italian, and

Mandarin speakers) tend to be exaggerated in vowel duration differences with spectral dimensions being less informative (Cebrian, 2006; Escudero, Benders, & Lipski, 2009; Flege, Bohn, & Jang, 1997). This can cause intelligibility problems for native listeners of English. In the present study, listeners were exposed to unfamiliar speech which sounds like foreign-accented English vowels deviating from English norms in their informativeness of a primary acoustic dimension. That is, when exposed to an uninformative primary acoustic dimension, listeners are expected to adapt to unfamiliar speech patterns by redirecting their attention to the most diagnostic acoustic dimension in speech sound categorization. We expect that perceptual adaptation may arise based on adjustment of acoustic-phonetic cues to phonological categories in which listeners redirect their attention to the most diagnostic acoustic dimension when a primary dimension becomes less useful in speech sound categorization.

2.1.2 Individual differences in perception of acoustic cues to speech

Although the majority of studies have focused on group-level differences in the perception of acoustic cues that define speech sound contrasts, a growing body of research has found large differences across individual listeners (e.g., Idemaru, Holt, & Seltman, 2012; Kapnoula, Winn, Kong, Edwards, & McMurray, 2017; Kong & Edwards, 2011; 2016). In particular, even though acoustic cues that contribute to category identity tend to be more strongly weighted than those less predictive of category identity, these acoustic cues are weighted differently across individual listeners (e.g., Beddor, 2015; Idemaru et al., 2012; Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). For example, Kong and Edwards (2011, 2016) examined perceptual cue weighting of VOT and f0 in the perception of the English stop voicing contrast and found that listeners differed considerably in the extent to which they use each acoustic dimension as a cue to the English stop voicing contrast. Beddor (2015) found that listeners differed in how much they use vowel nasalization cues in English and that individual differences in the use of vowel nasalization cues are also linked to nasalization of vowels in production. It has also been documented in previous studies that these individual differences in the perception of acoustic cues are stable over time (Idemaru et al., 2012; Kong & Edwards, 2016; Schertz, Cho, Lotto, & Warner, 2015; Yu & Lee, 2014).

While studies in individual differences in acoustic cue weighting have focused on whether cue weighting strategies differ across individuals and whether these differences are stable over time, how and to what extent listeners differ in adapting their use of multiple acoustic cues in response to unfamiliar pronunciations have received relatively less attention. Schertz et al. (2016) investigated differences in dimension-based statistical learning and whether they are related to individual differences in cue weighting strategies. Schertz et al. found that there is large individual variability in Korean listeners' cue weighting strategies for the English stop voicing contrast, and these initial cue weighting differences result in different patterns of adaptation during the Reverse block. That is, listeners who used VOT as a primary cue to the stop voicing contrast reduced their use of f0 as a secondary cue to the contrast whereas listeners who used f0 as a primary cue to the contrast reduced their use of VOT as a secondary cue. This indicates that this individual variability in cue weighting strategies is robust and it can provide the basis of listeners' adaptation strategies.

2.1.3 Cognitive abilities in speech perception

Recent studies have provided some evidence of potential sources of individual differences in speech perception. One possible source is cognitive abilities underlying speech perception processes. It has been suggested that general cognitive abilities such as working memory, attention, and inhibitory control aid more general learning processes (Goldstone, 1998). Previous research has pointed out a potential link between individual differences in general cognitive abilities and the perception of speech sounds (Akeroyd, 2008). Also, it has been observed that cognitive abilities contribute to individual performance on speech perception tasks even after controlling for auditory sensitivity (Füllgrabe, Moore, & Stone, 2015). To investigate contributions of cognitive abilities to speech perception processes, studies have tested a range of cognitive abilities as measured by executive functions, which refer to a set of cognitive processes that are needed for cognitive control of behavior when performing tasks and attaining goals (Diamond, 2013; Friedman & Miyake, 2017; Miyake & Friedman, 2012). In particular, three core executive functions have been suggested and extensively tested (Miyake & Friedman, 2012): inhibitory control, working memory, and cognitive flexibility. Inhibitory control (also known as inhibition) is the ability to suppress goal-irrelevant or competing information, and is commonly tested using psychological tests such as the Stroop task or the Flanker task (e.g., Bender, Filmer, Garner, Naughtin, & Dux, 2016). Working memory indicates the ability to hold information in the mind and simultaneously process it mentally. Working memory tasks include the Digit Span (forward or backward) task, the Corsi Block task, and the N-back task (e.g., Baddeley, 2003). Cognitive flexibility involves changing perspective or approaches to new rules or demands as in switching between tasks and is commonly tested using the Wisconsin Card Sorting task or the Trail-Making task (e.g., Kortte, Horner, & Windham, 2010).

These key components of executive function have been shown to account for some of the variance in speech perception in studies using a single test or a combination of executive function measures (Adank & Janse, 2010; Banks, Gowen, Munro, & Adank, 2015; Janse & Adank, 2012; Tamati, Gilbert, & Pisoni, 2013). For example, there is evidence that higher working memory capacity is associated with better speech perception abilities especially in speech perception in noise (Tamati et al., 2013). Also, some studies have shown that age-related differences in cognitive abilities may explain speech perception performance of older and younger adults (Adank & Janse, 2010; Janse & Adank, 2012). In Adank and Janse (2010), cognitive flexibility predicted differences in comprehension of a novel accent by younger and older adults. Inhibitory control has also been observed to be related to foreign-accent adaptation in older adults (Janse & Adank, 2012). Studies have also shown that certain cognitive abilities play an important role in individuals' adaptation to novel accents and unfamiliar speech (Banks et al., 2015). In Banks et al. (2015), for instance, individuals with better inhibitory control showed faster adaptation to the unfamiliar accent.

Despite these efforts in recent years, the exact role of cognitive abilities in speech perception processes has not been fully understood. That is, correlations between cognitive abilities and speech perception were generally weak or inconsistent in quite a few studies (Banks et al., 2015; Bent et al., 2016; Janse & Adank, 2012; Kim & Hazan, 2010; Kong & Edwards, 2016). For example, Kim and Hazan (2010) adopted several cognitive ability tasks such as inhibition, working memory, and attentional measures to examine whether cognitive abilities are related to individual differences in the learning of new speech contrasts. They found that a measure of attention switching was only weakly correlated with native English participants' ability to learn Korean stop contrasts. Bent et al. (2016) examined whether cognitive factors predicted individual differences in the perception of unfamiliar speech such as regional, nonnative, and disordered speech. Their results showed that listeners' vocabulary size was the only significant predictor of individual word recognition performance among the measures they used in the study including inhibitory control and cognitive flexibility. Similarly, Kong and Edwards (2016) found no significant relation between cognitive measures, such as inhibition and attention switching, and individual differences in gradiency in speech perception.

In addition to the core executive functions (i.e., inhibitory control, working memory, and cognitive flexibility), this study also examines sustained attention, which assesses individuals' ability to maintain attention for a certain amount of time (Jongman, Roelofs, & Meyer, 2015). This measure was included to control for the impact of general attentional maintenance on performance on the learning task. Overall, using a variety of cognitive measures, the current study aims to better understand the role of cognitive abilities in speech perception as to whether cognitive abilities contribute to better adaptation to unfamiliar speech patterns.

2.1.4 Categorization gradiency in speech perception

Although cognitive abilities may play a role in flexibility in speech perception, listeners' sensitivity to acoustic details also likely contributes to better adaptation to variability in speech. One such source of individual differences in speech perception is differences in phoneme categorization gradiency. Research has suggested that gradient encoding of speech categories, in which listeners are more sensitive to subtle acoustic differences such as within-category information, may require more flexible and efficient speech processing (Massaro & Cohen, 1983; Toscano, McMurray, Dennhardt, & Luck, 2010). These studies have postulated that gradient categorization behavior may be useful because it allows for flexibility in how acoustic cues are mapped onto sound categories.

Recently, several studies have shown that listeners vary in how gradient their categorization is (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016; Munson, Schellinger, & Edwards, 2017; Schellinger, Munson, & Edwards, 2017). As a measure of gradiency of phoneme categorization, these studies used a visual analog scaling (VAS) task, which is a continuous measure of phonetic categorization (Massaro & Cohen, 1983). Rather than forcing participants to choose between two options, participants are given a continuous line between two options and asked to mark their choice anywhere along the line. Studies using this task have found substantial individual differences. For example, in their study of the stop voicing contrast (/da/-/ta/), Kong and Edwards (2011, 2016) employed the VAS task and demonstrated that listeners differed significantly in their phoneme categorization responses. That is, some listeners exhibited a more categorical pattern in favor of endpoint responses while others showed a more gradient pattern using a wide range of available responses. In line with Kong and Edwards (2011, 2016), Kapnoula et al. (2017) also found that there are individual differences in phoneme categorization gradiency and importantly, that more gradient listeners' responses were also tightly linked to acoustic differences in the stimuli. In other words, for more gradient listeners, as the stimulus moved from one end of the continuum to the other, listeners responses shifted from one end of the scale to the other.

A second finding of these studies is that individuals who have more gradient categorization patterns are more sensitive to a secondary acoustic dimension (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). This suggests the possibility that these listeners would also be sensitive to changes in a secondary acoustic dimension. These studies have also found a trend that categorization gradiency in speech perception is associated with cognitive abilities although the trend is weak. However, the functional role of phoneme categorization gradiency remains to be fully understood.

The present study aims to confirm previous findings of individual differences in phoneme categorization gradiency, sensitivity to secondary cue use and the link to cognitive abilities and to extend them to the perception of vowels. Vowel perception has sometimes been described as more gradient than consonants in general, so the same patterns of individual variability may not hold. Furthermore, the study investigates whether gradiency predicts listeners' patterns of perceptual adaptation.

2.1.5 The present study

This study is primarily concerned with examining whether listeners flexibly adapt to unfamiliar speech sounds that deviate from long-term regularities of their native language by making short-term changes to acoustic cues. The unfamiliar speech sound pattern in the present study resembles those encountered in foreign-accented English vowels (e.g., Korean-, Italian-, or Mandarin-accented English vowels). In these cases, the relative informativeness of acoustic dimensions (spectral quality vs. duration) can be changed such that the most informative dimension (spectral quality) is no longer useful, but the role of the secondary cue (duration) is enhanced (Cebrian, 2006; Flege et al., 1997). More specifically, this study focuses on listeners' adaptive strategies to changes in the relative informativeness of acoustic dimensions such that most informative dimension becomes no longer useful. This study further investigates whether and to what extent individual differences in cognitive abilities and phoneme categorization gradiency are related to adaptation to these atypical speech sound categories. Research questions of this study are:

1. Do listeners flexibly adapt to unfamiliar speech that deviates from learned long-term regularities by increasing a secondary acoustic dimension when the most informative acoustic dimension is no longer diagnostic?
2. Are previously observed patterns of individual differences in phoneme categorization gradiency (i.e., more gradient vs. more categorical) also observed for vowel perception? If so, are patterns of categorization gradiency in vowels related to secondary cue use and cognitive abilities (i.e., inhibitory control, working memory, cognitive flexibility, and sustained attention)?
3. Do individual differences in phoneme categorization gradiency and cognitive abilities predict individual listeners' perceptual adaptability of phonetic categories?

We predict that listeners will increase their reliance on a secondary acoustic dimension when the most diagnostic dimension becomes no longer informative. We also expect considerable variability in the extent to which individuals adapt to unfamiliar speech. In terms of the potential sources of why individuals differ in their adaptation patterns, we assume that individual differences in cognitive abilities play a role in their adaptive patterns and thus we expect significant correlations of adaptation and cognitive abilities (Banks et al., 2015). In addition to cognitive ability measures which commonly assess core executive functions (i.e., inhibitory control, working memory, and cognitive flexibility), we also probe individuals' ability to maintain attention to control for the effect of listeners' maintenance of attention on task performance (Jongman et al., 2015). Based on previous findings on individual variability in phoneme categorization gradiency (Kapnoula et al., 2017; Kong & Edwards, 2016), we predict considerable individual differences which may be related to perceptual adaptability such that individuals who have a more gradient pattern of speech perception are more sensitive to secondary cues and in turn show better adaptation to unfamiliar speech sound categories.

2.2 Methods

2.2.1 Participants

Thirty-six monolingual speakers of Canadian English (mean age = 22, range = 18–31, 10 male) were paid for their participation. All participants reported normal hearing with no speech impairments.

2.2.2 Stimuli

Figure 2.1 illustrates stimuli for the VAS task and Baseline, Test and Exposure stimuli for the adaptation task. For the VAS and the adaptation stimuli, a female Canadian English talker from Ottawa recorded multiple utterances of *head* and *had* in a sound-proof booth with a high-quality recorder (Zoom H4n, 44.1 kHz sampling rate). The best tokens of *head* and *had* were then chosen and resynthesized to create a twenty-step continuum of spectral quality (from /ɛ/ to /æ/) using TANDEM-STRAIGHT in MATLAB (Kawahara, Takahashi, Morise, & Banno, 2009), which allows for making a natural-sounding spectral continuum from two natural end points. Eight native speakers of English were asked to identify *head* or *had* along the continuum and the most ambiguous step was chosen. Seven spectral steps were then chosen based on the most ambiguous token and the two end-point tokens. From each of the seven spectral steps, vowel duration continua ranging from 80 ms to 380 ms (50 ms/step) were created using the PSOLA algorithm in Praat (ver. 6.0.19, Boersma & Weenink, 2016). This procedure resulted in a total of 49 stimuli, orthogonally varying in two acoustic dimensions (7 steps formant frequencies × 7 steps vowel duration) from /ɛ/ to /æ/. An additional 12 stimuli were created in the same way from the same end-point recordings for the Exposure phase of the adaptation task (described below).

The adaptation stimuli consisted of the Baseline, Exposure and Test stimuli. The Baseline stimuli were a subset of the VAS stimuli, which included the full range of seven spectral steps at two different vowel duration steps (14 stimuli, repeated 7 times for a total of 98 trials). The Exposure stimuli consisted of 6 tokens of ambiguous formant frequencies and 12 adjacent ambiguous tokens to the most ambiguous tokens as shown in Figure 1 (18 stimuli, repeated 12 times for a total of 216 trials). The Baseline and the Exposure stimuli included the Test stimuli (red square and blue triangle). Comparison of responses to these spectrally ambiguous Test stimuli in each block assessed listeners' use of the secondary cue across the course of the experiment.

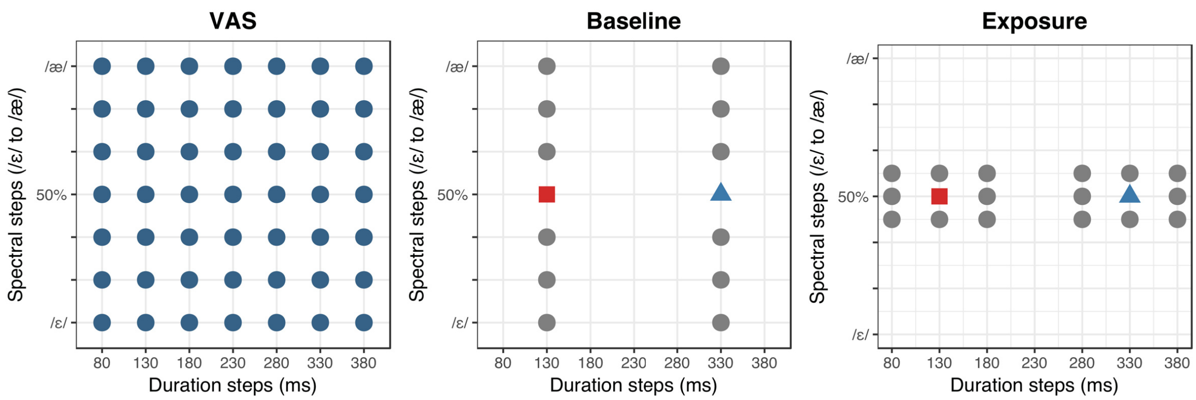

Figure 2.1. Illustration of the stimuli used in the VAS and the adaptation task. Test stimuli for the adaptation task were the red square and the blue triangle.

2.2.3 Procedure

Participants first completed the VAS task, followed by two cognitive tasks (i.e., Corsi and Berg Card Sorting Test), the adaptation task, and finally the other two cognitive tasks (i.e., Stroop and Continuous Performance Test). Participants sat in front of a computer and were tested individually in a sound-attenuated booth after receiving both oral and written instructions about the experiments. The experiments were conducted at McGill University, Canada.

2.2.3.1 The VAS task

The VAS task was administered before the adaptation task (a two-alternative forced choice identification; 2AFC) to minimize any step-like bias induced by the 2AFC task on the VAS task (Kapnoula et al., 2017). In the VAS task, each participant heard 245 trials of 7 spectral × 7 duration continuum (5 repetitions) randomly using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). Upon hearing each stimulus, a double-headed arrow was displayed on the computer monitor. One end of the arrow was labeled as *head* and the other end was labeled as *had*, and participants were instructed to click a location along the line that corresponded with the percept of proximity to *head* or *had*. The VAS task was completed in approximately 17 minutes.

2.2.3.2 The cognitive tasks

Three subsets from the Psychology Experiment Building Language (PEBL, Mueller & Piper, 2014) were administered to assess major components of executive functions: the Stroop Color and Word Test (Stroop), the Corsi block-tapping test (Corsi), and the Berg Card Sorting Test (BCST). Additionally, one attentional measure from PEBL was administered to assess sustained attention, which indicates the maintenance of vigilance and one of the primary components of attention (Cohen, 2014): the Continuous Performance Test (CPT). The Stroop task is a measure of inhibitory control in which participants see the names of colors (e.g., green) in colored text (e.g., blue) and respond to the color of the text, not the word itself, by pressing the corresponding key (MacLeod, 1991). In the compatible condition the color of the text and the word match (e.g., the word green in green text), and in the incompatible condition the color of the text and the word mismatch (e.g., the word red in green text). The Corsi task is a measure of working memory (Vandierendonck, Kemps, Fastame, & Szmalec, 2004). On each trial, participants see an array of blocks and are shown a sequence of highlighted blocks, starting with a sequence of two blocks and gradually increasing in length up to nine blocks. Participants must then click on the blocks with the mouse in the same sequence. The BCST is a computerized version of the Wisconsin Card Sorting Test in PEBL, which is a measure of cognitive flexibility (Miyake et al., 2000). In this task, participants classify cards according to one of three classification rules (i.e., color, shape, or number), which change every 10 cards. Participants receive feedback as to whether they applied the rule correctly or not. Participants must figure out the changing rules, and the task measures how well they adapted to the changing rules. The CPT is a measure of sustained attention (Conners, Epstein, Angold, & Klaric, 2003). In this task, participants responded to a constant series of letter stimuli on the computer screen and responded to all stimuli except the letter 'x' for approximately 14 minutes. The cognitive tasks in total took approximately 30 minutes.

2.2.3.3 The adaptation task

The adaptation stimuli were presented as a 2AFC task in MATLAB, in which listeners heard the words *head* and *had* and identified the word they heard with a key press. The Baseline block was presented first in which each participant heard 98 trials of 7 spectral × 2 duration steps (7 repetitions). This was followed by the Exposure block in which each participant heard 216 trials of 3 spectral × 6 duration steps (12 repetitions). Participants also repeated the same Baseline block (as in Baseline 1 and Baseline 2) after the Exposure block. All the trials within a block were randomly presented through headphones at a comfortable listening level. The adaptation task was completed in approximately 25 minutes.

2.2.4 Analysis

2.2.4.1 The VAS task

The analysis of the VAS task closely followed prior work (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). The click location for each trial was measured in pixels. The monitor screen was 1280 × 800 pixels in size. Click locations on the x-axis were converted to a VAS rating scale (1–100) based on Kapnoula et al. (2017). Clicks that were more than 3 standard deviations away from the y-axis mean (391 observations, 4.4% of data) were removed. In order to quantify degree of gradiency, for each individual, a rotated logistic function was fit following Kapnoula et al. (2017).[1] Gradiency was assessed using the slope of the rotated logistic function (shallower slopes (smaller values) indicate more gradient responses).

As in previous studies (Kapnoula et al., 2017; Kong & Edwards, 2016), the effect of secondary cue use on categorization gradiency was investigated using differences in crossover point for the two continua (short and long vowel) from the baseline 2AFC task. Crossover points were measured for each participant by fitting a four-parameter logistic function and using the midpoint variable (see Kapnoula et al., 2017 for details). The crossover differences offer a measure of secondary cue use (i.e., multiple cue integration) that is independent of the VAS task (Kapnoula et al., 2017; McMurray, Samelson, Lee, & Tomblin,

[1] The rotated logistic fits two parameters. Theta is the angle of diagonal boundary line in the two-dimensional space defined by the two cues, and is assumed to reflect relative use of the two cues. Then a logistic curve is fit orthogonal to this boundary and the estimated slope of this curve is used as a measure of gradiency, independent of cue use.

2010).

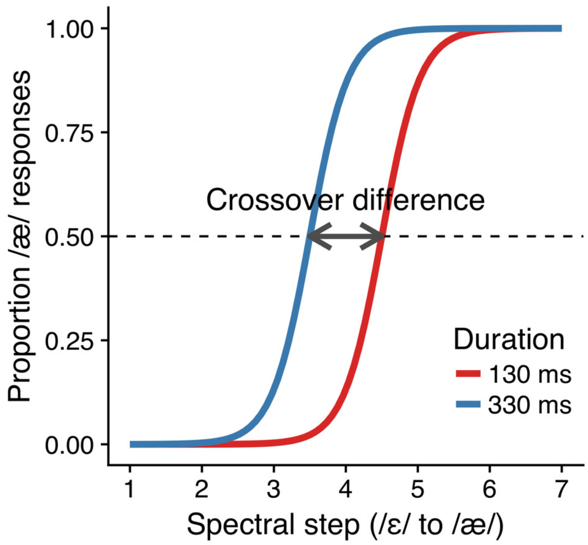

Figure 2.2. Hypothetical illustration of duration cue use at Baseline 1 as measured by the difference in 2AFC crossover points between short and long vowel durations.

2.2.4.2 Cognitive measures

Individual inhibitory control performance was assessed by Stroop interference—the average difference between response time in incongruent and neutral trials in milliseconds (MacLeod, 1991). A higher Stroop interference value corresponds to poorer inhibitory control. Individual working memory performance was recorded as the total Corsi task score, which was defined as the correct sequence in the correct serial location (Vandierendonck et al., 2004). A higher Corsi task score indicates better working memory capacity. For cognitive flexibility, total perseverative errors of the BCST were calculated for individual listeners (Fox, Mueller, Gray, Raber, & Piper, 2013). More perseverative errors on the BCST indicate less cognitive flexibility. Also, individual sustained attention performance was assessed based on proportion target accuracy of the CPT (Conners et al., 2003). More accurate responses on the CPT reflect better sustained attention. Based on these cognitive task measures, a correlation analysis will be conducted to examine whether cognitive abilities are correlated with one another across individuals. After examining correlations between cognitive tasks, they will be entered as predictors in a multiple linear regression analysis of performance on the VAS task.

2.2.4.3 The adaptation task

Perceptual adaptation will be measured in terms of significant changes in listeners' categorization responses to Test stimuli from Baseline 1 to Exposure. This study also examines whether listeners adapt back to canonical pronunciations from Exposure to Baseline 2 when they hear canonical exemplars of their native language at Baseline 2. A mixed-effects logistic regression analysis will be used to investigate whether individual listeners' responses are predicted by individual difference measures of phoneme categorization gradiency, secondary cue use and cognitive ability measures (inhibitory control, working memory, cognitive flexibility, and sustained attention). Statistical models will be described more in detail in the results section.

2.3 Results

2.3.1 Vowel categorization at Baseline 1

Figure 2.3 shows vowel categorization at Baseline 1 for short and long vowel durations. The overall pattern of categorization responses indicates that listeners mostly use spectral differences to categorize the vowel contrast. There was also an effect of vowel duration in their categorization responses but to a much weaker degree as expected (Escudero, 2000; Hillenbrand et al., 2000; Kondaurova & Francis, 2008; 2010; Liu & Holt, 2015). A mixed-effects logistic regression analysis with random intercepts and random slopes for spectral and duration steps for participants confirmed that listeners primarily rely on vowel spectral quality ($\beta = 10.167$, $z = 13.721$, $p < 0.001$) although vowel duration also contributes to vowel categorization ($\beta = 0.405$, $z = 7.505$, $p < 0.001$). The results indicate a unique contribution of each acoustic dimension to vowel categorization responses after controlling for each other and reflect native English listeners' long-term representations of this vowel contrast. It should be noted that Figure 2.3 also indicates considerable individual differences in the use of vowel duration for vowel categorization, which will be discussed in relation to phoneme categorization gradiency in Section 3.3.

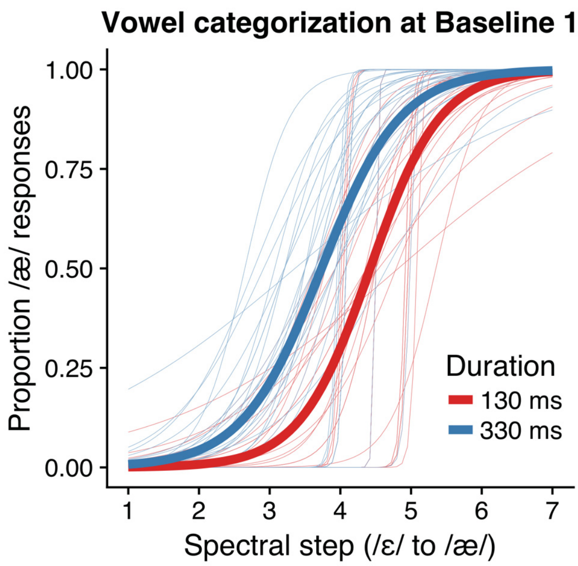

Figure 2.3. Proportion of /æ/ responses along vowel spectral quality continuum at Baseline 1 as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are logistic curves fit to each individual listener data for each vowel duration.

2.3.2 VAS and cognitive measures

2.3.2.1 The VAS task

Figure 2.4 shows VAS responses averaged across all participants. Overall, listeners used the entire line when making their responses although they responded more using the two endpoints of the line. This might indicate that some listeners made more categorical response patterns while other made more gradient response patterns.

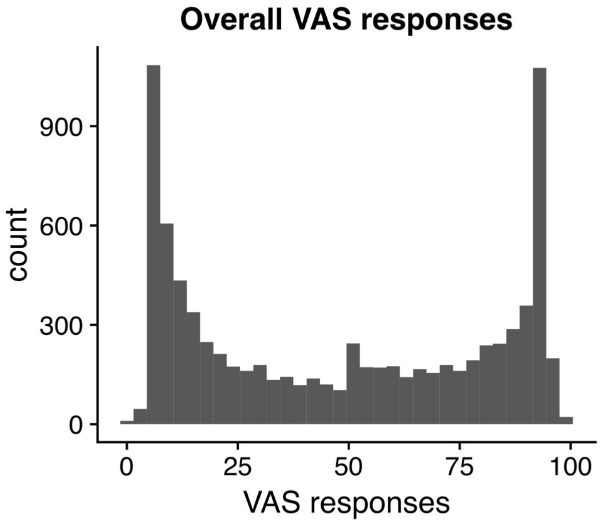

Figure 2.4. A histogram of overall visual analog scaling (VAS) responses.

Figure 2.5 shows results for three representative participants who made more categorical responses (Participant 187), less categorical responses (Participant 192), and more gradient responses (Participant 188). In Figure 2.5, participants' responses were illustrated by plotting overall VAS responses using histograms (top), VAS responses as a function of vowel spectral quality (middle) and duration (bottom). Specifically, the panels of Participant 187 show that VAS responses were largely categorical clustered around the two endpoints (top) and the responses were variable at the category boundary as a function of spectral quality (middle) while response patterns were mostly random as a function of duration (bottom). In contrast, the panels of Participant 188 show that VAS responses were more distributed across the entire line (top) and the responses shifted systematically as a function of spectral quality (middle) and responses were closer to the middle of the VAS scale (bottom), which shows a quite different pattern of responses from categorical listeners (e.g., Participant 187).

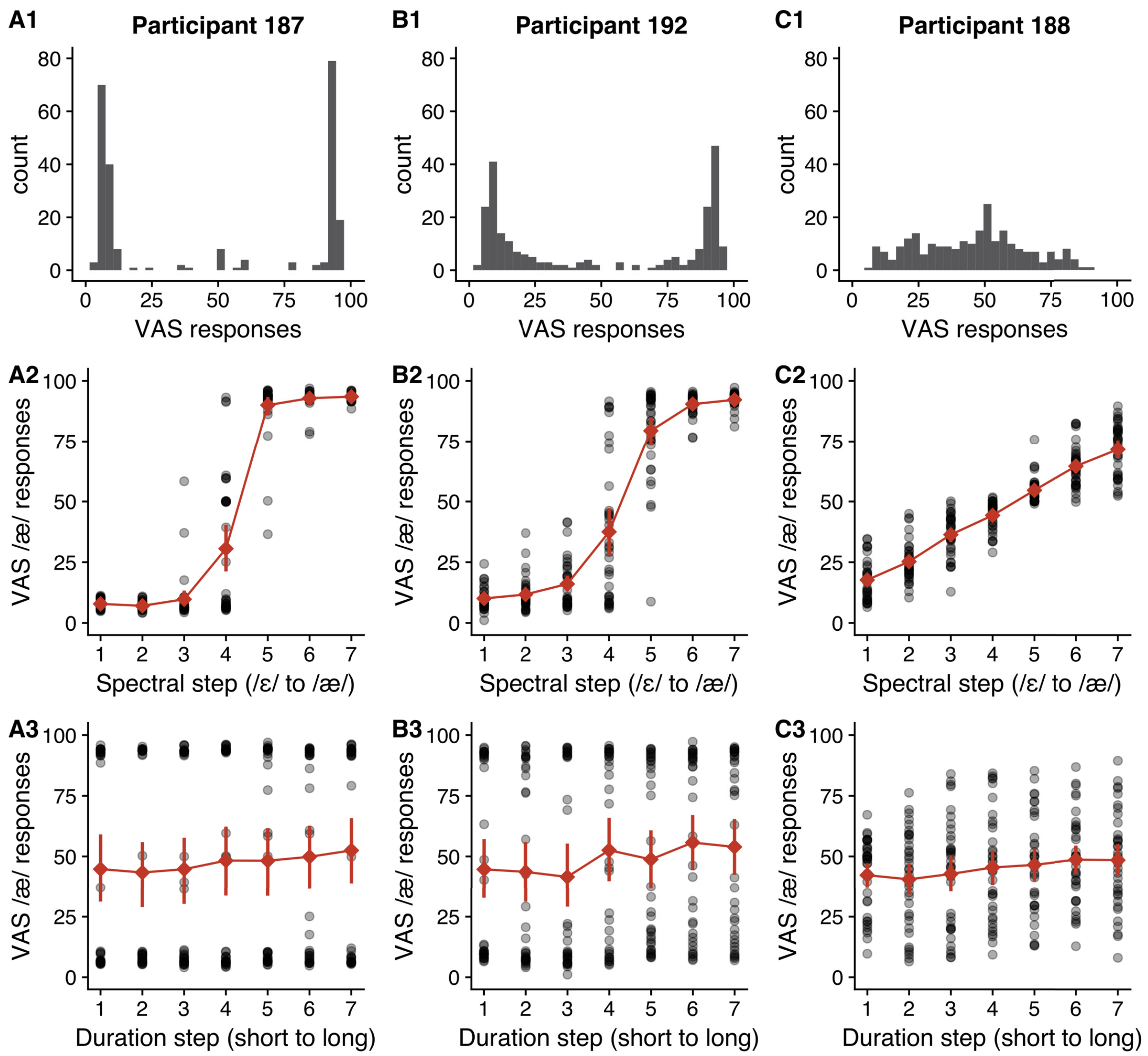

Figure 2.5. Visual analog scaling (VAS) responses for three representative listeners (more categorical vs. more gradient). The VAS slope values (gradiency measure) of each representative listener are 132 (Participant 187), 55 (Participant 192), and 10 (Participant 188) in which smaller values indicate more gradient responses.

Together, the results from the VAS task suggest that there were substantial individual differences in gradiency in phoneme categorization. Some listeners showed more categorical responses using the two end points while others showed more gradient responses using the entire line. These differences are quantified and compared to other measures in the following sections.

2.3.2.2 Relationship between individual difference measures

Before including individual difference measures in a statistical model, a correlation analysis was conducted to examine whether they are correlated to one another. Table 2.1 shows the correlation matrix between individual difference measures. The four cognitive measures were Stroop interference (STROOP; inhibitory control), Corsi scores (CORSI; working memory), BCST task perseverative errors (BCST; cognitive flexibility), and CPT task accuracy (CPT; sustained attention). The correlation analysis also included gradiency (VAS; gradiency) and the crossover difference between two duration steps at baseline (CODIFF; secondary cue use). Among all individual difference measures only VAS and CORSI were significantly correlated ($r = –0.50$, $p < 0.01$), indicating that gradient responses are linked to better working memory capacity. This also indicates that each cognitive measure may tap into a different cognitive ability. These cognitive ability measures were subsequently included in a linear regression analysis along with CODIFF, to examine whether secondary cue use affects categorization gradiency.

Table 2.1. Correlation matrix between cognitive ability measures.

	STROOP	BCST	CORSI	CPT	VAS	CODIFF
STROOP	—					
BCST	–0.19	—				
CORSI	0.13	–0.17	—			
CPT	0.07	–0.17	–0.05	—		
VAS	–0.04	0.30	–0.50**	–0.17	—	
CODIFF	0.03	0.06	–0.16	–0.21	0.19	—

(**$p < 0.01$)

2.3.2.3 Relationship between VAS and other measures

In order to analyze the contribution of secondary cue use and cognitive abilities to categorization gradiency, a multiple linear regression analysis was conducted. All measures were continuous and they were standardized by centering and dividing by 2 standard deviations before they were entered into the model. Table 2.2 shows the results of the regression model for categorization gradiency. Each coefficient is the estimated effect when all other predictors are controlled for.

Table 2.2. Summary of linear regression model predicting categorization gradiency. Model coefficient estimates (β), standard errors (SE), corresponding *t*-values, and *p*-values.

Predictor	Estimate (β)	SE	*t*	*p*
Intercept	37.654	3.378	11.147	< 0.001
CoDiff	–16.759	7.126	–2.352	0.025
Stroop	5.002	7.037	0.711	0.482
BCST	9.963	7.163	1.391	0.174
Corsi	–27.235	7.131	–3.819	< 0.001
CPT	–11.931	7.167	–1.665	0.106

Figure 2.6 shows two significant predictors of the regression model. In the model, CoDiff (secondary cue use) significantly predicted the VAS slopes (phoneme categorization gradiency) (β = –16.759, t = –2.352, p = 0.025), as shown in Figure 2.6A. This indicates that listeners who use the secondary cue more also gave more gradient responses in phoneme categorization. This is consistent with previous findings in which the use of a secondary cue predicts gradiency in phoneme categorization (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). The regression model also yielded a significant relation between Corsi scores and VAS slopes (β = –27.235, t = –3.819, p < 0.001), as shown in 2.6B. That is, individuals with higher working memory capacity also made more gradient responses in phoneme categorization, in line with Kapnoula et al. (2017).

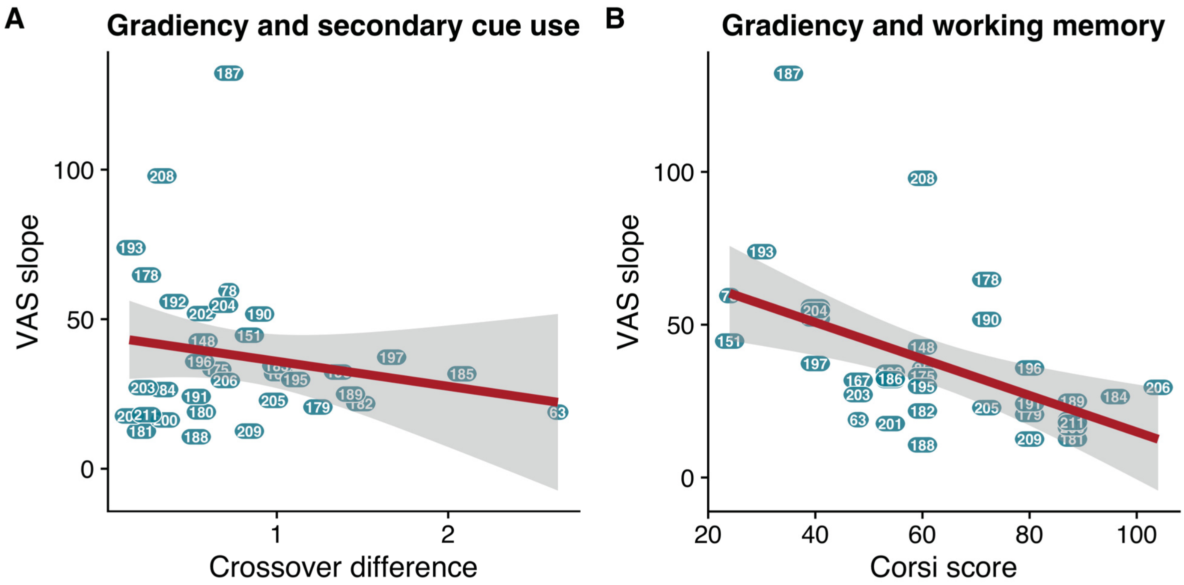

Figure 2.6. Categorization gradiency as a function of secondary cue use and working memory. The shallower VAS slopes (smaller values) indicate more gradient responses. Higher Corsi scores indicate better working memory capacity.

2.3.3 The adaptation task

This section presents the results of the adaptation task and whether patterns of adaptation are associated with individual difference measures described above (i.e., cognitive ability measures and categorization gradiency). To briefly recap, it was hypothesized that listeners would flexibly adapt to unfamiliar pronunciations (e.g., tokens in the Exposure block of the adaptation task) by showing an increased reliance on a secondary dimension (i.e., vowel duration) when the most informative dimension is not diagnostic (i.e., uninformative spectral quality in Exposure vs. informative spectral quality in Baseline 1 and Baseline 2). It was also hypothesized that variability in the extent to which individuals adapt to unfamiliar speech would be predicted by cognitive and speech processing differences across individuals as measured by cognitive ability tasks and phoneme categorization gradiency, respectively.

To examine the adaptability of categorization responses, the participants' proportion of /æ/ responses to Test stimuli were analyzed using mixed-effects logistic regression, using the *glmer()* function from the *lme4* package (ver.1.1-16) in R (R Core Team, 2017). A mixed-effect logistic regression model was built to examine the effects of phoneme categorization gradiency and cognitive abilities on perceptual adaptation. All continuous variables—VAS slopes (VAS; gradiency), Stroop interference effects (STROOP; inhibitory control), Corsi scores (CORSI; working memory), BCST task perseverative errors (BCST; cognitive flexibility), and CPT task accuracy (CPT; sustained attention)—were standardized by centering and dividing by 2 standard deviations. DURATION was centered (–0.5 and 0.5) and examined changes in the use of vowel durations to adapt to non-canonical speech patterns across experimental blocks. BLOCK was coded using sum contrasts comparing Baseline 1 and Exposure ($Block_1$) and also Baseline 1 and Baseline 2 ($Block_2$), to examine whether listeners' categorization responses change from Baseline 1 to Exposure and return at Baseline 2, respectively. The model included by-participant random intercepts and by-participant random slopes for BLOCK, DURATION, and their interaction.

Table 2.3. Summary of perceptual adaptation model. Model coefficient estimates (β), standard errors (SE), corresponding z-values, and p-values.

Predictor	Estimate (β)	SE	z	p
Intercept	–0.605	0.138	–4.390	< 0.001
BLOCK_1 (Baseline 1 vs. Exposure)	–1.292	0.181	–7.135	< 0.001
BLOCK_2 (Baseline 1 vs. Baseline 2)	0.632	0.196	3.216	< 0.001
DURATION	–2.220	0.143	–15.490	< 0.001
VAS	–0.917	0.352	–2.600	0.009
STROOP	1.031	0.290	3.551	< 0.001
BCST	0.276	0.288	0.959	0.337
CORSI	0.340	0.317	1.074	0.282
CPT	–0.432	0.275	–1.568	0.116
BLOCK_1 × DURATION	1.357	0.431	3.147	0.001
BLOCK_2 × DURATION	–0.743	0.396	–1.876	0.060
BLOCK_1 × VAS	–0.273	0.499	–0.549	0.583
BLOCK_2 × VAS	–0.698	0.526	–1.327	0.184
BLOCK_1 × STROOP	0.270	0.385	0.702	0.482
BLOCK_2 × STROOP	0.623	0.431	1.443	0.148
BLOCK_1 × BCST	–0.623	0.362	–1.720	0.085
BLOCK_2 × BCST	0.565	0.371	1.522	0.128
BLOCK_1 × CORSI	–0.476	0.380	–1.252	0.210
BLOCK_2 × CORSI	–0.379	0.416	–0.912	0.361
BLOCK_1 × CPT	–0.723	0.311	–2.325	0.020
BLOCK_2 × CPT	0.240	0.345	0.695	0.486
DURATION × VAS	–1.832	0.396	–4.622	< 0.001
DURATION × STROOP	0.123	0.311	0.396	0.692
DURATION × BCST	–0.183	0.271	–0.676	0.499
DURATION × CORSI	0.291	0.294	0.990	0.322
DURATION × CPT	–1.195	0.243	–4.911	< 0.001
BLOCK_1 × DURATION × VAS	–0.639	1.178	–0.542	0.587
BLOCK_2 × DURATION × VAS	–0.308	1.053	–0.293	0.769
BLOCK_1 × DURATION × STROOP	1.778	0.910	1.952	0.050
BLOCK_2 × DURATION × STROOP	–1.475	0.868	–1.699	0.089
BLOCK_1 × DURATION × BCST	0.532	0.887	0.600	0.548
BLOCK_2 × DURATION × BCST	–0.965	0.746	–1.293	0.195
BLOCK_1 × DURATION × CORSI	–0.400	0.953	–0.420	0.674
BLOCK_2 × DURATION × CORSI	0.177	0.834	0.213	0.831
BLOCK_1 × DURATION × CPT	–0.023	0.812	–0.029	0.977
BLOCK_2 × DURATION × CPT	–0.219	0.694	–0.136	0.751

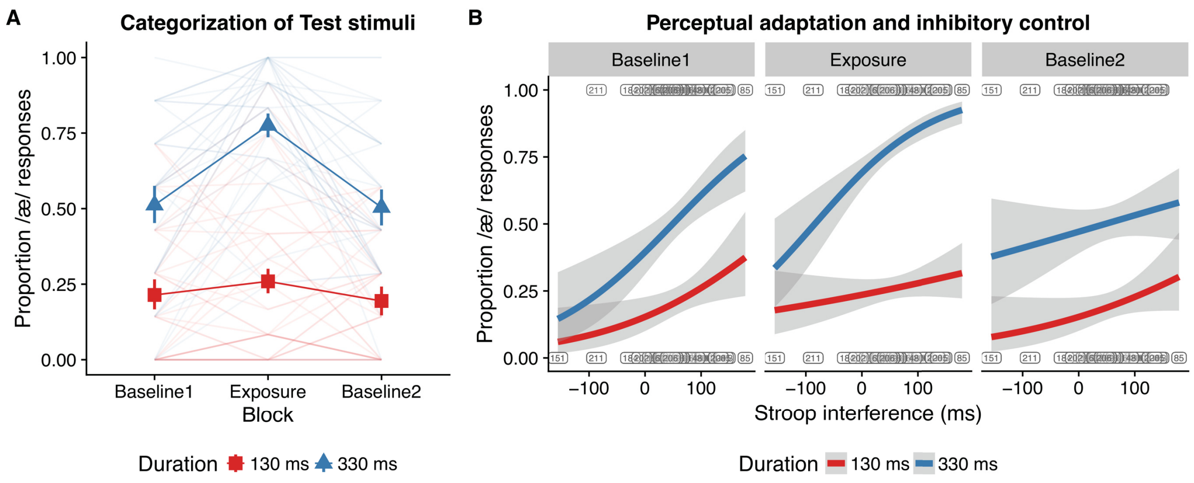

Figure 2.7. (A) Proportion of /æ/ vowel responses of the Test stimuli across blocks as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are individual listeners' Test stimulus responses across blocks. (B) The effect of individual differences in Stroop interference (inhibitory control) on categorization responses for short (130 ms) and long vowel (330 ms) durations across blocks. A higher Stroop interference score indicates low inhibitory control, and each number indicates individual listeners.

Although the model included all lower terms relevant to three-way interactions involving BLOCK, DURATION and VAS, and also BLOCK, DURATION and cognitive measures (i.e., STROOP, CORSI, BCST, CPT), we will focus on the two-way interaction involving BLOCK and DURATION to investigate perceptual adaptation and three-way interactions involving BLOCK and DURATION and the individual difference measures (VAS and cognitive measures) to investigate whether they predicted perceptual adaptation.

Figure 2.7A shows proportion of /æ/ vowel responses of the Test stimuli across blocks as a function of short and long vowel durations, and Figure 2.7B shows how individual differences in Stroop interference (inhibitory control) influence vowel categorization responses for short and long vowel durations across blocks. As shown in Figure 2.7A, there was a two-way interaction of BLOCK_1 × DURATION, indicating that the difference between short and long durations is bigger at Exposure than that at Baseline 1 ($\beta = 1.357$, $z = 3.147$, $p = 0.001$). This suggests that listeners exhibited a significant up-weighting of reliance on the duration dimension in the Exposure block when the spectral dimension was not informative for vowel categorization. Figure 2.7A also shows that listeners overall flexibly down-weighted their use of duration at Baseline 2 when they heard speech input which is consistent with the long-term English norm although there is a marginal two-way interaction of $\text{B}\textsc{lock}_2 \times \text{D}\textsc{uration}$ ($\beta = -0.743$, $z = -1.876$, $p = 0.060$), indicating a subtle difference in responses between Baseline 1 and Baseline 2.

Individual listeners' Test stimulus responses across blocks (as indicated in thin lines in Figure 2.7A) illustrate considerable individual variability in up-weighting of the duration dimension in the Exposure block. Accordingly, three-way interactions in the model investigate factors that could predict a substantial amount of the individual variability in the adaptation performance. The model found a significant three-way interaction of $\text{B}\textsc{lock}_1 \times \text{D}\textsc{uration} \times \text{S}\textsc{troop}$ ($\beta = 1.778$, $z = 1.952$, $p = 0.050$), indicating that greater perceptual adaptation at Exposure is associated with poorer inhibitory control (Figure 2.7B). This issue will be addressed in the discussion section in terms of poor inhibitory control as a broader focus of attention.

Although we found no evidence of a relationship between gradiency and adaptation, nor between other cognitive abilities (i.e., working memory, cognitive flexibility, and sustained attention) and adaptation, the two-way interaction involving CPT (sustained attention) is worth mentioning. That is, the model found a significant two-way interaction of $\text{B}\textsc{lock}_1 \times \text{CPT}$ ($\beta = -0.723$, $z = -2.325$, $p = 0.020$) and of $\text{D}\textsc{uration} \times \text{CPT}$ ($\beta = -1.195$, $z = -4.911$, $p < 0.001$). These interactions indicate that individuals with higher sustained attention are more sensitive to varying experimental conditions across blocks and duration differences of the input signal, respectively, relative to individuals with lower sustained attention. However, because there was no 3-way interaction between CPT, DURATION and $\text{B}\textsc{lock}_1$ sustained attention was not linked to the magnitude of duration change (i.e. to the magnitude of perceptual adaptation).

2.4 Discussion

The current study examined perceptual adaptability of speech sounds categories when confronted with changes in the informativeness of cues in the input signal. More specifically, we found that listeners adapted their cue weighting strategies to the unfamiliar vowels by up-weighting a secondary acoustic dimension (i.e., duration) when they were exposed to an ambiguous primary acoustic dimension (i.e., spectral quality). We also found considerable variability in the extent to which individuals adapt to unfamiliar speech, and that this variability was related to individual differences in cognitive abilities (i.e., inhibitory control).

2.4.1 Perceptual adaptation to unfamiliar speech

The current results confirmed previous findings (Idemaru & Holt, 2011; 2014; Liu & Holt, 2015) that listeners exhibited dynamic adaptation to unfamiliar phonetic categories in which they initially adapt to unfamiliar speech patterns at Exposure and subsequently switched their representations back to their long-term category representations when they heard canonical English pronunciations at Baseline 2. The current results further suggest that the speech perceptual system adjusts to the acoustic consequences of changes in the *relative informativeness* of acoustic dimensions. That is, after only brief exposure to unfamiliar speech patterns listeners increased their reliance on a secondary acoustic dimension to maintain a phonetic contrast when a primary dimension becomes no longer informative and a secondary dimension was the only reliable information available for phonetic categorization.

Crucially, the current adaptation task differs from that in previous research on dimension-based perceptual learning in that listeners adapted to atypical phonetic categories by up-weighting a secondary dimension rather than down-weighting it. The current task also differs from that in the previous work in that listeners had to increase reliance on a secondary dimension to adapt without recourse to the most reliable anchor dimension signaling category membership (i.e., a primary dimension), which was always available in the previous work. Nevertheless, both of these results suggest that listeners dynamically adapt to short-term deviations in the input signal while simultaneously maintaining stable long-term representations. Up-weighting a secondary cue in perceptual adaptation can be interpreted as a compensatory strategy of secondary cue enhancement to adapt to adverse listening conditions. This type of compensatory strategy, at least for vowels, has also been observed in speech production to improve intelligibility (Ferguson & Kewley-Port, 2007; Schertz, 2013). For example, Schertz (2013) found that speakers exaggerated duration differences between the segments in English when they clarified misheard speech, especially for tense and lax vowels. This secondary cue enhancement in speech production was also reported in Ferguson and Kewley-Port (2007) in which speakers increased vowel duration differences to improve vowel intelligibility in clear speech compared to conversational speech. These findings from speech production suggest that enhancing secondary cues may be a common compensatory strategy in both speech perception and speech production.

The listeners' adaptive strategy in the current study where listeners up-weighted a secondary dimension when the most diagnostic dimension became uninformative indicates that listeners are sensitive to the distributional properties of the input speech signal and adjust acoustic cues differently as a function of their informativeness in distinguishing phonetic categories (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Toscano & McMurray, 2010). Importantly, the present study showed that without any explicit training listeners can shift their attention to a more informative acoustic dimension as they adapt to unfamiliar pronunciations (cf. Francis & Nusbaum, 2002; Francis, Baldwin, & Nusbaum, 2000). Shifting of cue weights to reflect informative dimensions may be a mechanism that is used under other circumstances as well. Azadpour and Balaban (2015) examined the mechanisms underlying perceptual adaptation to spectrally-distorted speech (i.e., spectrally-rotated speech) by comparing phoneme category remapping, inverse transformation of spectral rotation, and changes in cue weighting strategies. They found that only changes in cue weighting strategies (i.e., shifting attention from spectral information to temporally-dynamic information) predicted perceptual adaptation to spectrally distorted speech. That is, listeners gave more weight to the acoustic information in the signal that was least affected by the distortion, which is also most reliable in making phonetic category decisions.

2.4.2 Relationship between gradiency and secondary cue use

The present results confirmed previous findings that individual differences in categorization gradiency are associated with secondary cue use in such a way that more gradient listeners showed greater use of a secondary cue (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). This finding suggests that listeners who show a gradient pattern are more sensitive to fine-grained acoustic information and thus are better at utilizing subtle acoustic differences across multiple cues. This also relates to the cue integration account in previous work in which multiple cue integration was linked to efficient sensory processing (Franken et al., 2017; Kapnoula et al., 2017). For example, using multimodal speech perception such as auditory and visual cues, Franken et al. (2017) found that individuals integrated auditory and visual information to re-adjust vowel categories and pointed out that listeners with less sharp category boundaries assigned more weight to a secondary cue (i.e., visual information) during audiovisual speech perception.

2.4.3 The role of categorization gradiency in perceptual adaptation

In line with previous studies, we found that individual listeners differed considerably in how gradiently they perceive speech sounds, but we found no evidence of a link between a more gradient pattern of phonetic categorization and more adaptation to atypical speech patterns. This might be related to the research design of this study in which learning involves more use of the secondary cue, and therefore more gradient listeners with more secondary cue use at Baseline have less room to make changes in the Exposure block. Alternatively, it is possible that perceptual adaptation processes lie not at the encoding level but at the decision level, in which some gradient listeners were sensitive to the secondary dimensions but decided not to use them to adapt to the unfamiliar patterns because they might have thought that it was not a diagnostic dimension in vowel categorization. This might indicate that both gradient and categorical listeners might be similarly adaptive for solving the problem of perceptual adaption. That is, although listeners exhibit differences in categorization gradiency, different degrees of gradiency may not contribute to solving this adaptation problem.

2.4.4 The role of cognitive abilities in perceptual adaptation

This study demonstrated that individual differences in inhibitory control predicted perceptual adaptation to unfamiliar speech. In the present study, inhibitory control was the only measure that was significantly correlated with the patterns of perceptual adaptation to unfamiliar speech sound categories. This finding is in accordance with previous observations of the close link between inhibitory control and speech perception (Darcy, Mora, & Daidone, 2016; Lev-Ari & Peperkamp, 2013; 2014). Upon close inspection, however, the present result indicates that individuals with poor inhibitory control showed better adaptation to unfamiliar speech by enhancing a secondary dimension to adapt. This finding might be surprising if one assumes that the ability to suppress goal-irrelevant information is beneficial in most contexts. It is possible that individuals who show strong inhibitory control are more likely to selectively attend to goal-relevant information and in turn are more likely to miss all other information. In the case of vowel categorization in English, the most relevant information is vowel spectral quality and vowel duration is less relevant. This might indicate that vowel duration is less likely to be available for individuals who exert more inhibitory control. On the other hand, vowel duration may be more readily available for individuals who exert less inhibitory control as they may have a broader focus of attention, subsequently facilitating their ability to adapt to unfamiliar speech using secondary cues.

Regarding incidental benefits of poor inhibitory control, recent studies have suggested that reduced inhibitory control can enhance learning performance under some circumstances (Amer, Anderson, & Hasher, 2018; Amer, Campbell, & Hasher, 2016). These studies have shown that reduced inhibitory control (e.g., less likely to involve active suppression of irrelevant information) may lead to a broader focus of attention and process more information, which may be beneficial in certain contexts. Similarly, in their face recognition study Weeks and colleagues (Weeks, Biss, Murphy, & Hasher, 2016) found that when participants were shown faces with a name that they were instructed to ignore, individuals with poor inhibition (i.e., older adults in their study) performed better at associating faces with corresponding names. They interpreted this finding as an indication that reduced control of suppressing task-irrelevant information may be beneficial in some learning contexts which depend on utilizing less goal-relevant information. If these previous findings are also pertinent to the current result, individuals with low inhibitory control might have been at an advantage relative to individuals with high inhibitory control on the adaptation task in which less relevant information in vowel categorization (i.e., duration) suddenly became relevant and listeners were required to learn less relevant information to adapt. However, more work is needed to establish the relationship between inhibitory control and adaptation processes in speech.

Because the ability to maintain attention should be beneficial to learning activities (Wickens & McCarley, 2008), some might wonder whether the amount of attention participants paid during the task is attributable to recognition of speech sounds or perceptual adaptation. In the present study, we tested individual listeners' ability to maintain alertness over time to control for the effect of sustained attention ability on adaptation performance, as well as other executive function measures. The finding of this study showed that individual differences in sustained attention is not predictive of the patterns of perceptual adaptation in speech. The study also showed that categorization gradiency in speech perception is not related to listeners' ability to maintain their attention during the task.

2.4.5 Working memory for speech and adaptation

The current finding showed that higher working memory capacity is linked to more gradient processing of speech sounds. This is in line with the result of Kapnoula et al. (2017) in which higher working memory capacity was associated with phoneme categorization gradiency in the VAS task. This may indicate that higher working memory capacity benefits gradient speech perception by facilitating processing and retention of fine-grained within-category differences. On the other hand, perceptual adaptation was not related to individual differences in working memory. Rather, adaptability was linked to another cognitive ability (i.e., inhibitory control). These findings suggest that different cognitive abilities may underlie gradiency and adaptation in speech perception. That is, adaptation to unfamiliar speech may also involve other perceptual processes at the decision level as well as processing at the encoding level. Non-significant correlations between cognitive ability measures provide additional evidence that these factors may differentially affect listeners' perceptual responses.

2.5 Conclusion

The findings of the present work add to a growing body of research addressing how individual cognitive and perceptual abilities influence speech perception (Banks et al., 2015; Bent et al., 2016; Kapnoula et al., 2017; Kong & Edwards, 2016). This study further confirms previous findings that there are considerable individual differences in the perception and perceptual adaptability of speech sound categories (Kapnoula et al., 2017; Kong & Edwards, 2016; Schertz et al., 2016). Specifically, the present results suggest that listeners are sensitive to distributional information of the speech input and flexibly adapt to unfamiliar speech sound categories using a more informative acoustic dimension while simultaneously maintaining stability of their long-term phonetic representations of their native language. The present results also suggest that there are considerable individual differences in their adaptation patterns and these differences are in part accounted for by individual differences in inhibitory control. That is, better adaptation to non-native-like English vowels was related to lower inhibitory control (i.e., a broader focus of attention), which was beneficial when less relevant information became useful in this case. Together, this study provides insights into the interplay between speech and cognitive processes and contributes to a better understanding of the mechanisms underlying flexibility in speech perception.

Preface to Chapter 3

Chapter 2 examined perceptual adaptability of speech sounds categories when confronted with changes in the informativeness of cues in the input signal. The findings of Chapter 2 indicate that listeners flexibly adapt their cue weighting strategies to unfamiliar pronunciations which simulate non-native English vowels by increasing their use of a secondary acoustic dimension when a primary dimension becomes no longer diagnostic. We also found considerable differences in the extent to which individuals adjust their cue weighting strategies, and these differences were linked to individual differences in inhibitory control. Chapter 3 continues to investigate flexibility in speech perception in challenging listening conditions. Compared to Chapter 2 which addressed talker-related challenges (i.e., unfamiliar speech), Chapter 3 examines an adverse condition involving receiver limitations (i.e., cognitive load) and aims to simulate a real-life multi-tasking condition which commonly involves concurrent auditory and visual tasks. The goal of this chapter is to better understand the mechanisms underlying plasticity in speech perception when listeners cope with cognitive load caused by dual-task situations and what makes some listeners better adapters to the demands of distracting conditions.

Chapter 3

Study 2: Individual differences in flexibility in speech perception under cognitive load

3.1 Introduction

In everyday life, listeners commonly hear speech sounds under sub-optimal listening conditions. It is in fact rarely the case that listeners hear speech in fully focused and quiet listening conditions as in the laboratory setting. Often times, understanding speech is part of a multi-task in which it coincides with other tasks, for example, hearing a lecture while taking notes or hearing an address of the location you are looking for while searching for a road sign. In these multi-tasking situations, listeners generally have more difficulty in understanding speech and their comprehension of speech tends to be impaired likely due to increased cognitive load (Mattys, Davis, Bradlow, & Scott, 2012 for a review). Although listeners may have reduced comprehension of speech under these challenging conditions, they may also use adaptive strategies to compensate for their increased cognitive load and reduced attentional resources, achieving good understanding of speech. Study 1 in the previous chapter addressed how listeners adjust their cue weighting strategies for phonetic categories to overcome a talker-related challenge (i.e., foreign-accented speech). Study 2 also addresses adaptive cue weighting strategies that listeners may utilize to overcome adverse conditions but a challenge of a different type (i.e., listener-related) to gain insight into adaptive plasticity in speech perception under different adverse conditions. More specifically, the goal of Study 2 is to better understand how and to what extent listeners' speech perception abilities are modulated by cognitive load when listeners are engaged in a dual task and what makes some listeners better adapters to the demands of distracting conditions.

3.1.1 Speech perception in adverse conditions

To perceive speech sounds and recognize the meaning, listeners constantly overcome a wide variety of challenging conditions including talker-related challenges (e.g., accented, non-native, and atypical speech), environmental degradation (e.g., noise and background babble), and listeners' limited processing resources (e.g., cognitive load and divided attention) (cf. Mattys et al., 2012). Understanding how listeners overcome adverse listening conditions is a fundamental issue for better understanding the perceptual and cognitive mechanisms underlying speech perception and speech communication more generally. A considerable number of studies have investigated speech perception in a variety of adverse conditions (e.g., Bradlow & Bent, 2008; Cainer, James, & Rajan, 2008; Mattys et al., 2012). For example, studies have shown that relatively brief exposure to foreign-accented speech can lead to intelligibility improvements for the accented talker (Bradlow & Bent, 2008; Clarke & Garrett, 2004). Such intelligibility improvements were also found after encountering speech in noise (Cainer et al., 2008). In addition to improvements in intelligibility, several studies have also shown that listeners adapt to accented speech by using top-down lexical knowledge (Norris et al., 2003). Thus, listeners have been shown to overcome these challenges by flexibly adapting to these varying situations and achieving perceptual constancy (e.g., Guediche, Blumstein, Fiez, & Holt, 2014; Mattys et al., 2012; Norris, McQueen, & Cutler, 2003; Zhang & Samuel, 2014) though this may come at the cost of increased effort (McGarrigle, Munro, Dawes, Stewart, Moore, Barry, & Amitay, 2014).

3.1.2 The effect of cognitive load on speech perception

Among the aforementioned adverse conditions, the current study pays particular attention to the effects on speech perception of cognitive load generated by a dual task (e.g., Mattys & Palmer, 2015; Mattys & Wiget, 2011; Mattys, Barden, & Samuel, 2014; Mitterer & Mattys, 2017). Cognitive load caused by dual-task situations is quite common in everyday listening conditions (e.g., multi-tasking) but have received relatively little attention compared to talker-related challenges such as foreign-accented speech (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004) and environmental degradation such as background noise (e.g., Broersma & Scharenborg, 2010; Tamati, Gilbert, & Pisoni, 2013). However, studying the effects of cognitive load on speech perception is important because it involves cognitive processes such as the control of attentional and memory resources and allow us to understand the role that those processes play in speech perception (Mattys et al., 2012).

It has been suggested that speech perception is an inherently attention demanding process (Magnuson & Nusbaum, 2007) and that limited attentional resources typically have detrimental impacts on speech perception (e.g., Mattys & Palmer, 2015; Mattys & Wiget, 2011; Mitterer & Mattys, 2017). For example, using a dual-task paradigm in which listeners carried out a visual search task while hearing speech, Mattys and Wiget (2011) tested the acuity of speech perception under a baseline no load condition and a cognitive load condition in which listeners performed a concurrent visual search task while perceiving speech. They found that discrimination performance on pairs of syllables on the /g/-/k/ continuum (e.g., *gi*-*ki*) was poorer in the cognitive load than the no load condition, indicating that cognitive load generally impairs the perception of acoustic details in the signal. Subsequent work further suggested that this potentially disruptive effect of cognitive load on speech perception occurs at the early stages of speech perception—the encoding of the speech signal in memory (Mattys et al., 2014; Mitterer & Mattys, 2017).

Cognitive load has also been shown to impact the relative weighting of primary and secondary acoustic cues in speech perception (Gordon, Eberhardt, & Rueckl, 1993; Kong & Lee, 2017). Each phonological contrast is signaled by multiple acoustic-phonetic cues (Lisker, 1986). Studies have shown that listeners give more attention to some acoustic cues over others, and the relative importance of these cues in phonetic categorization has been termed cue weighting (Francis, Kaganovich, & Driscoll-Huber, 2008; Holt & Lotto, 2006). Using a similar dual-task paradigm (e.g., categorizing speech sounds while solving math problems) as the previous studies Gordon et al. (1993) showed that under load listeners' cue weights decreased overall, but that reliance on primary acoustic cues decreased more relative to secondary cues with cognitive load. This in turn led to a modest increase in the effect of the secondary cues on categorization responses under cognitive load. This suggests that the greater weights given to primary cues (e.g., VOT as a cue to the English stop voicing contrast) in optimal listening conditions may be attenuated to a larger extent than those given to secondary cues (e.g., f0 as a cue to the English stop voicing contrast) in adverse conditions. This disruptive effect of cognitive load on acoustic cues can be interpreted as noisy encoding as in Gordon et al. (1993) in which a poorer signal-to-noise ratio leads to poorer encoding. Thus, it is expected that cognitive load would generally exert a detrimental effect on the encoding of acoustic cues.

At the individual listener level, however, Kong and Lee (2017) showed substantial differences across individuals in the magnitude of attentional modulation on cue weighting under cognitive load. In particular, quite a few listeners in fact revealed increased cue weights under cognitive load although this pattern did not receive particular attention in their discussion. This may be surprising based solely on the noisy encoding account in Gordon et al. (1993), in which the processing acuity of acoustic cues is likely to be impaired under cognitive load. However, Mattys and Wiget (2011) found that while load affected discrimination scores it did not affect the steepness of categorization functions—often taken as an indicator of acuity (e.g., Hazan, Messaoud-Galusi, Rosen, Nouwens, & Shakespeare, 2009). Furthermore, when individual differences in performance were re-analyzed and described in Mitterer and Mattys (2017), there was no overall pattern at the individual level, though they noted a trend for steeper slopes under cognitive load. This might indicate that individual listeners differ in the extent to which their speech categorization patterns are modulated in the distracting condition, with some showing steeper slopes and others shallower. In other words, although at the group level the robustness of cue weighting has been shown to be attenuated and use of secondary cues has been observed to be boosted in previous studies (Gordon et al., 1993; Kong & Lee, 2017), it is also possible that different cue weighting strategies may be masked at the individual level. For example, it is possible that some listeners may increase their reliance on a primary cue or a secondary cue as a compensatory strategy and others have more difficulty adapting to the distracting condition and only show attenuation of cue weights without using any compensatory strategies.

The present study explores whether and how individuals differ in their use of multiple acoustic cues under cognitive load. Understanding the nature of and factors contributing to such individual differences is crucial for better understanding the mechanisms underlying speech perception processes and their potential interplay with other listener characteristics such as individual cognitive abilities. In particular, speech perception in challenging conditions is an ideal ground for testing individual differences in speech perception. In addition to differences in how individuals' cue weighting strategies are modulated under cognitive load, the present study investigates potential sources of individual differences in speech perception, relating adjustment of cue weighting strategies under cognitive load to listeners' cognitive abilities and sensitivity to fine-grained acoustic details, which will be described in the following sections.

3.1.3 The role of cognitive abilities in speech perception

An emerging body of research in recent years has highlighted the role of cognitive abilities in speech perception especially in adverse conditions (e.g., Banks, Gowen, Munro, & Adank, 2015; Bent, Baese-Berk, Borrie, & McKee, 2016; Janse & Adank, 2012; Tamati et al., 2013).

In the present study, we focus on two key components of cognitive abilities in relation to speech perception under cognitive load, i.e., working memory and inhibitory control. First, working memory indicates the ability to hold information in the mind and simultaneously process it mentally (Baddeley, 2003) and has been linked to speech perception performance in poor listening conditions in a considerable body of research (e.g., Akeroyd, 2008; Heald & Nusbaum, 2014; Tamati et al., 2013; Wingfield, 2016). For example, there is evidence that higher working memory is associated with better speech perception abilities especially in speech perception in noise (Gordon-Salant & Cole, 2016; Tamati et al., 2013). Similarly, studies have also found a link between working memory and listeners' ability to understand degraded speech (Akeroyd, 2008; Besser, Koelewijn, Zekveld, Kramer, & Festen, 2013). However, although there is considerable evidence that working memory is important for speech perception in background noise as in the aforementioned studies, how working memory contributes to the perception of speech in distracting conditions remains to be fully understood.

In dual-task conditions, in particular, the interference of a secondary task (e.g., a concurrent visual search task) on a primary speech task (e.g., speech categorization) would increase working memory demands, resulting in a high degree of cognitive load (Mitterer & Mattys, 2017). We predict that although overall speech perception performance declines under distracting conditions, individuals with high working memory capacity may perform better. This is because high working memory capacity may contribute to the ability to keep more information relevant to speech perception in working memory which may compensate for increased processing demands under cognitive load (Sörqvist & Rönnberg, 2014). We predict that individuals with high working memory capacity may retain information better than those with low working memory capacity and thus are less likely to show poorer performance under cognitive load.

In addition to working memory capacity, inhibitory control has also been observed to be associated with speech perception in adverse conditions (Banks et al., 2015; Janse, 2012). Inhibitory control (also known as inhibition) is the ability to suppress goal-irrelevant or competing information (Diamond, 2013). More specifically in speech perception, inhibition may play a role in attentional processes that are involved in actively suppressing irrelevant acoustic information or selectively attending to more relevant acoustic information. There are a few studies that examined the link between individual differences in inhibitory control and speech perception processes (Banks et al., 2015; Janse, 2012; Tamati et al., 2013). For example, Janse (2012) found that inhibitory control was associated with speech perception performance on degraded speech in older adults, but this pattern was not found in young adults (Tamati et al., 2013). Banks et al. (2015) found that individuals with better inhibitory control showed faster adaptation to the unfamiliar accent, but inhibition did not significantly correlate with overall speech perception performance. In contrast, in Study 1 in the previous chapter, we found that weaker inhibitory control is linked to better adaptability of unfamiliar vowel categories. Thus, despite a growing body of research on the link between inhibitory control and the perception and adaptation of speech sounds, the results have been mixed, and little is known about how inhibitory control is related to speech perception under cognitive load. In the present study, we expect that inhibitory control may allow listeners to better overcome distracting listening conditions due to their cognitive control to attend to more relevant information by ignoring distractions.

It should be noted that the studies described above have mostly relied on one cognitive task for a measure of each cognitive construct (e.g., using the backward digit span task to examine working memory). Recent studies on individual differences in cognitive abilities have suggested that individual tasks designed to target a particular cognitive construct are not always strongly correlated with each other, likely due to task-specific effects (Friedman & Miyake, 2017; Hedge, Powell, & Sumner, 2017; Schmiedek, Lövdén, & Lindenberger, 2014). Thus, any given cognitive task may not reliably reflect a cognitive construct on its own, rather combining the results of multiple tasks of the same cognitive construct is necessary to obtain more reliable measures underlying cognitive abilities. The present study uses multiple measures of working memory and inhibition to obtain more reliable measures of individual differences in these abilities.

3.1.4 The role of gradiency in speech perception

Another relevant factor that may contribute to individual variability in speech perception is gradiency in phoneme categorization. Studies have observed that listeners are sensitive to within-category differences and they perceive speech sounds in a gradient fashion (Kapnoula et al., 2017; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008). These studies have suggested that gradiency in phoneme categorization is maintained to the level of lexical activation. Recent studies have further observed that individual listeners vary substantially in how much they show gradiency when they categorize speech sounds (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016; Munson et al., 2017; Schellinger et al., 2017). Using a visual analog scaling (VAS) task as a measure of phoneme categorization gradiency, these studies have also shown that individuals who had more gradient categorization patterns were more sensitive to secondary acoustic cues. These studies have examined individual listeners' weighting of two cues (VOT and f0) in the perception of the stop voicing contrast in English and have found that more gradient listeners are more sensitive to a secondary acoustic dimension (f0) (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). They also found that categorization gradiency is not consistently related to inhibition (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016) but is associated with working memory (Kapnoula et al., 2017). In Chapter 2, we also found that phoneme categorization gradiency was predicted by working memory capacity. These findings are particularly relevant to the current study because categorization gradiency may not be closely linked to speech perception performance in optimal conditions, but may be linked to speech perception in adverse listening conditions with respect to use of detailed within-category differences. Adopting the VAS task as a measure of gradiency in phoneme categorization, the present study investigates whether individuals differ in their patterns of phoneme categorization and whether the potential variability of speech perception gradiency correlates with listeners' changes in cue weighting strategies under cognitive load.

3.1.5 The present study

The overarching goal of this study is to shed light on the mechanisms underlying speech perception by better understanding individual differences in speech perception under cognitive load. The current study first investigates whether listeners' cue weighting strategies are modulated under cognitive load using a dual-task paradigm in which listeners perform a visual search task while perceiving speech sounds (Bosker, Reinisch, & Sjerps, 2017; Mattys & Palmer, 2015; Mattys & Wiget, 2011; Mitterer & Mattys, 2017). We predict that listeners' cue weighting strategies would be modulated with reduced attentional resources and cognitive load would interfere more with primary cues than secondary cues (Gordon et al., 1993; Kong & Lee, 2017).

This study also examines whether individual listeners differ in patterns of perceptual cue weighting strategies under cognitive load. Crucially, it is expected that listeners would differ widely in the extent to which their use of multiple acoustic cues is modulated with attentional demands. That is, although the cue weights given to primary cues may be overall mitigated more than those given to secondary cues, the extent to which cue weighting strategies are influenced by attentional disruption may differ across individuals. For example, some listeners may increase their reliance on a primary acoustic cue as a compensatory strategy while other listeners may increase their reliance on a secondary acoustic cue under cognitive load. Still others would be less likely to use these compensatory strategies under conditions of high cognitive load. This study tests two different cognitive sources of these expected individual differences: working memory and inhibitory control. First, we expect that higher working memory would protect speech perception performance such that listeners with higher working memory capacity would be less affected under divided attention than listeners with lower working memory capacity. That is, perceptual cue weights would be overall less likely to be impaired or may even be enhanced for listeners with higher working memory capacity to compensate for limited attentional resources. We also expect that listeners with good inhibitory control would be less affected under divided attention. That is, their use of the primary cue in particular may be less affected due to their better attention to the most relevant information. Listeners with good inhibitory control may also be more likely to increase the use of a relevant primary dimension as a compensatory strategy under divided attention. Due to potential task-specific effects of cognitive tasks (Friedman & Miyake, 2017; Hedge et al., 2017; Schmiedek et al., 2014), the present study uses multiple measures of the same cognitive construct to obtain more reliable measures of individual differences in cognitive abilities. Finally, we test the hypothesis that individual differences in listeners' phoneme categorization gradiency (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016) are associated with the effects of attentional demands on the use of multiple acoustic cues. We predict that gradient listeners, who have been shown to be more sensitive to fine acoustic details, will be more likely to adapt to the distracting condition by showing compensatory strategies such as increasing primary and/or secondary cues.

3.2 Methods

3.2.1 Participants

Fifty-seven speakers of North American English were paid for their participation. Of these, 3 participants were excluded from further analysis because they did not complete all tasks (2) or were not a native speaker of North American English (1). All the remaining participants (*N*

= 54, mean age = 22, range = 18–33, 21 male) reported normal hearing with no speech impairments.

3.2.2 Stimuli

Figure 3.1 shows stimuli for the VAS task and auditory stimuli for the dual task. In preparation for the stimuli, a female Canadian English speaker from Ottawa in her 20s first recorded multiple utterances of *heed*, *hid*, *head*, and *had* in a sound-proof booth with a high-quality recorder (Zoom H4n, 44.1 kHz sample rate). The best tokens of *heed* and *hid* for the VAS task and those of *head* and *had* for the dual task were chosen and resynthesized to create a 20 step continuum of vowel spectral quality for each vowel contrast (i.e., /i/-/ɪ/ for the VAS task and /ɛ/-/æ/ for the dual task) using TANDEM-STRAIGHT in MATLAB (Kawahara et al., 2009). From these 20 steps for each continuum, seven spectral steps for the VAS task and five spectral steps for the dual task were then chosen based on the most ambiguous token and two natural end points for each continuum through a preliminary test. From each of the seven spectral steps of *heed* and *hid*, vowel duration continua ranging from 80 ms to 350 ms (45 ms/step) were created using the PSOLA algorithm in Praat (ver. 6.0.03, Boersma & Weenick, 2015). The same procedure was applied to each of the five spectral steps of *head* and *had* to create duration continua ranging from 120 ms to 360 ms (60 ms/step). The 5 step continuum of *head* and *had* rather than 7 step continuum was chosen to minimize the length of the dual task.

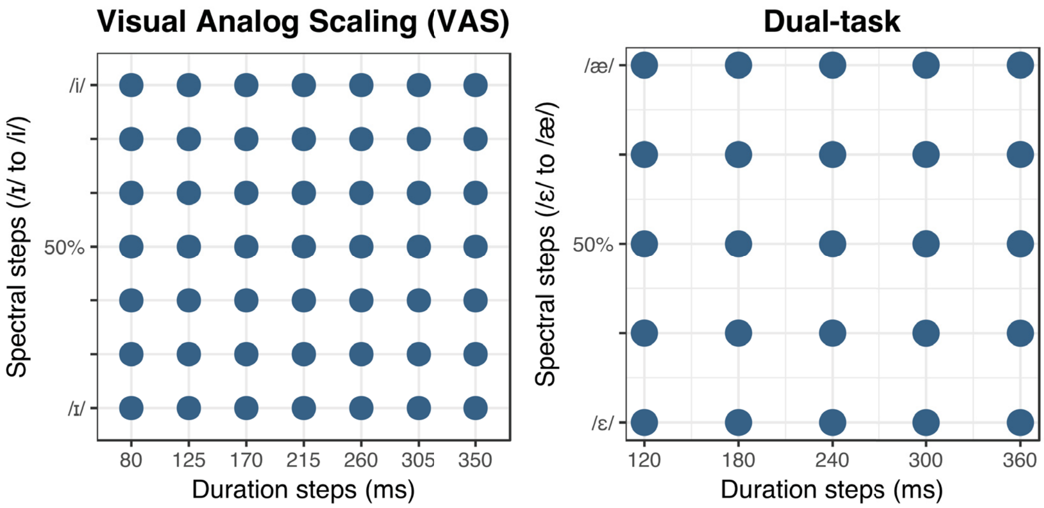

Figure 3.1. Illustration of the stimuli in the VAS and the auditory stimuli in the dual task.

Figure 3.2 shows visual stimuli for the dual task, modeled after previous work (Bosker et al., 2017; Mattys et al., 2014; Mattys & Wiget, 2011; Mitterer & Mattys, 2017). The stimuli consisted of odd-one-out arrays (13 × 13) of colored shapes in which about half of the arrays (i.e., 13 out of 25 arrays) contain a black diamond (the target) adapted from Bosker et al. (2017).

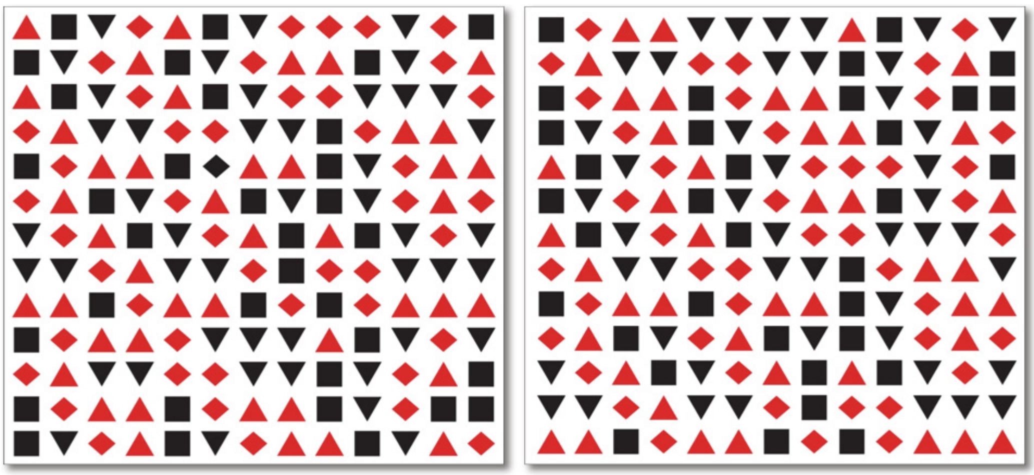

Figure 3.2. Example stimuli for the concurrent visual search task. The left panel contains an oddball object (black diamond) and the right panel does not.

3.2.3 Procedure

Participants first completed the visual analog scaling (VAS) task, then one working memory and one inhibition task (i.e., the backward digit span task and the Stroop task (Stroop Color and Word Test)), followed by the dual task (i.e., two alternative forced choice with a concurrent visual search), and another working memory and inhibition task (i.e., the reading span task and the Go/No-go task). Participants sat in front of a computer and were tested individually in a sound-attenuated booth after receiving both oral and written instructions about the experiments. The experiments were conducted in a quiet room at McGill University, Canada.

3.2.3.1 The VAS task

In the VAS task, each participant heard 196 trials of 49 stimuli items (7 spectral × 7 duration continuum) repeated 4 times in random order using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Upon hearing each stimulus, participants saw a double-headed arrow on the computer monitor. One end of the arrow was labeled as *heed* and the other end was labeled as *hid*. Participants were instructed to click a location on the line based on their judgment of how close the stimulus was to either *heed* or *hid*. The VAS task was completed in approximately 15 minutes.

3.2.3.2 The dual task

In the dual task, participants first completed a baseline two-alternative forced choice (2AFC) identification task (the baseline no load condition) and then 2AFC with a concurrent visual search task (the cognitive load condition) in which participants judged whether a black diamond was present in an array of colored shapes. The baseline no load condition was a 2AFC task in which participants heard one of the stimuli and identified whether they heard *head* or *had* with a key press. Each participant heard 125 trials of 25 stimuli items (5 spectral × 5 duration continuum) repeated 5 times in random order. After having a self-paced short break, participants proceeded with the cognitive load condition. On each trial of the cognitive load condition, a fixation dot was displayed in the middle of the screen for 500 ms, immediately followed by an object grid which was displayed for 4000 ms. An audio stimulus was presented with a 1500 ms delay after the onset of the object grid to make sure that participants heard the audio stimulus while they were engaging fully in visual search. After the end of the visual display, participants were asked to identify the word they heard (either *head* or *had*) with a key press and then whether or not they had seen a black diamond (the oddball object). Each participant heard 125 trials of 25 audio stimuli (5 spectral × 5 duration steps) repeated 5 times in random order as in the no load condition, and also saw 125 trials of 25 visual stimuli (13 object grids with a diamond and 12 object grids without a diamond) repeated 5 times in random order. All stimuli were presented in OpenSesame (Mathôt et al., 2012), and the dual task including both conditions was completed in approximately 25 minutes.

3.2.3.3 The cognitive ability tasks

Participants performed four cognitive tasks: two working memory tasks—the backward digit span task and the reading span task, and two inhibitory control tasks—the Stroop task and the Go/No-go task. The two shorter cognitive tasks (i.e., the backward digit span task and the Stroop task) were administered before the dual task to minimize fatigue of the participants before the dual task. The remaining two tasks (i.e., the Go/No-go task and the reading span task) were then administered after the dual task. All tasks were from the Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014) and took approximately 25 minutes.

In the backward digit span task (Dobbs & Rule, 1989), participants were presented with a series of digits (e.g., 5-7-8-2) and had to recall them in reverse order by typing in their responses with a computer keyboard (e.g., 2-8-7-5). List length increased from two three-digit trials to two eight-digit trials (i.e., two trials for each list length, 12 trials in total). The task ended when participants made two consecutive mistakes for a specific digit length. In the reading span task (Unsworth, Redick, Heitz, Broadway, & Engle, 2009), participants were asked to read a sentence, to determine whether it is logical, and to memorize a letter presented after the judgment. List length varied from three three-digit trials to three seven-digit trials after completing two two-digit trials (i.e., two trials for two-digit and three trials for two- to seven-digit trials, 17 trials in total). After each series of trials, participants were asked to recall letters in the correct order by typing in their responses with a computer keyboard. In the Stroop task (MacLeod, 1991), participants saw colored words and responded to the color, not the meaning by pressing the corresponding key. In the congruent condition the color of the word and the meaning are the same (e.g., the word green in green colored text), and in the incongruent condition the color of the word and the meaning are different (e.g., the word green in red colored text). In the Go/No-go task (Bezdjian, Baker, Lozano, & Raine, 2009), participants were asked to click the mouse button to rapidly respond to one stimulus category (Go stimuli, e.g., the letter ‘P’) while refraining from responding to the other (No-go stimuli, e.g., the letter ‘R’).

3.2.4 Analysis

3.2.4.1 Cognitive ability measures

The backward digit span task was scored on the total number of correctly recalled digit sequences in reverse order for each participant (Huettig & Janse, 2016). The reading span task was scored on the total number of correctly recalled letters in the correct position for each participant (Nagaraj & Magimairaj, 2017). For both working memory tasks, a higher score corresponds to better working memory capacity. The Stroop task performance was assessed based on the interference effect; that is, response time (RT) to incongruent trials – RT to neutral trials in milliseconds (MacLeod, 1991). For the Go/No-go task, the mean RT for Go trials and for No-go trial errors (commission errors) were measured for each participant. For both inhibitory control tasks, lower inhibition values correspond to better inhibitory control.

To obtain a more general working memory measure underlying the two memory measures and a more general inhibitory control measure underlying the two inhibition measures (Friedman & Miyake, 2017; Miyake, Emerson, & Friedman, 2000a), composite scores of working memory and inhibitory control were calculated by converting each cognitive task score to a z-score and aggregating (Bezdjian et al., 2009). However, as will be shown in the results section, the two inhibition tasks were not significantly correlated and thus a composite score of inhibitory control was not used in the analysis. Instead, separate Stroop and Go/No-go scores were used in the analysis. These cognitive ability measures were used in a mixed-effects logistic regression analysis to examine whether individual differences in working memory and inhibitory control measures are related to changes in individual cue weighting strategies under cognitive load.

3.2.4.2 The VAS task

The analysis of the VAS task was modeled after previous studies (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). Monitor size was 1680 × 1050, and the click location for each trial was measured in pixels. All responses that were more than ±40 pixels from the line in the y-axis (2513 observations, 26% of data) were excluded from the analysis. Click locations on the x-axis were converted to a VAS rating scale (1–100) for each trial (Kapnoula et al., 2017). Each individual's responses were averaged over the five repetitions for each stimulus. These responses were then fit by a rotated logistic function (Kapnoula et al., 2017) and gradiency was measured using the slope of the rotated logistic. In this measure, shallower slopes indicate more gradient responses. As with the cognitive ability measures, this individual gradiency measure was used in a mixed-effects logistic regression analysis to examine whether individual differences in gradiency in speech perception are associated with modulation in individual cue weighting strategies under cognitive load.

3.2.4.3 The dual task

The categorical responses from the 2AFC task were analyzed by fitting a mixed-effects logistic regression model using the *glmer()* function from the *lme4* package in R (R Core Team, 2017). The listeners' categorization responses were used as a dependent variable and fixed effects variables in the model included spectral step (SPECTRUM), duration step (DURATION), the cognitive load condition (CONDITION), cognitive measures (STROOP; inhibition, GO/NO-GO; inhibition, and WM; working memory composite scores) and VAS slopes (VAS; gradiency). The interaction terms between acoustic dimensions (SPECTRUM and DURATION) and the cognitive load condition (CONDITION) were included in the model to examine whether listeners' cue weighting strategies change under cognitive load. The interaction terms between acoustic measures, the cognitive load condition, and individual difference measures (cognitive abilities and categorization gradiency) were also included to examine whether the effect of cognitive load on acoustic dimensions is modulated by individual differences in cognitive abilities and gradiency. CONDITION was centered (–0.5 load and 0.5 no load) and all continuous variables, namely SPECTRUM, DURATION, STROOP, GO/NO-GO, WM, and VAS were standardized by centering and dividing by 2 standard deviations. The model included by-participant random intercepts and random slopes for SPECTRUM, DURATION, CONDITION and the interaction of CONDITION with the two acoustic dimensions.

Individual listeners' cue weights to each acoustic dimension for each load condition as a separate model were estimated by adding the random coefficients to the fixed effect coefficient of each acoustic cue and then were extracted from the random slopes for each cue for each listener (Kong & Edwards, 2016; Kong & Lee, 2017). Each individual's change in cue weights under cognitive load was calculated by subtracting the random slope coefficients of the cue in the no load condition from the cognitive load condition (Kong & Lee, 2017). These cue weight differences across conditions were subsequently used to assess how performance on the visual search task is related to performance on the speech perception task and were also used in figures involving the effect of cognitive load on perceptual cue weighting at the individual level to demonstrate how cue weight differences across conditions are linked to cognitive abilities and speech perception gradiency.

3.3 Results

3.3.1 Individual difference measures

3.3.1.1 Cognitive measures

Correlation analyses were conducted to examine whether cognitive measures are related to one other and whether using composite scores underlying different working memory and inhibitory control tasks is justifiable. Table 3.1 shows the correlation matrix between cognitive ability measures. Results of the Pearson correlation showed a significant positive correlation between the two working memory measures ($r(52) = 0.44$, $p < 0.001$), i.e., backward digit span scores (DSPAN) and reading span scores (RSPAN). These working memory measures were then converted to z-scores and were aggregated to obtain a working memory composite score (WM). Unlike working memory measures, the two inhibition measures—Stroop interference effects (STROOP) and Go/No-go scores (GO/NO-GO)—were not significantly correlated ($r(52) = 0.06$, $p = 0.66$). This suggests that the Stroop task and the

Go/No-go task may tap into different components of inhibitory control and using a composite inhibition score by combining Stroop and Go/No-go scores may not be justifiable. Thus, in the present study separate Stroop and Go/No-go scores were used rather than using one composite score for inhibitory control. These cognitive measures (i.e., WM; working memory composite scores, STROOP; Stroop interference effects, and GO/NO-GO; Go/No-go scores) were subsequently included in a mixed-effects logistic regression analysis to determine whether cognitive abilities are linked to changes in cue weighting strategies under cognitive load.

Table 3.1. Correlation matrix between cognitive ability measures.[2]

	STROOP	GO/NO-GO	DSPAN	RSPAN
STROOP	—			
GO/NO-GO	0.06	—		
DSPAN	–0.17	–0.29*	—	
RSPAN	–0.20	–0.40**	0.45***	—

(***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

3.3.1.2 Speech perception gradiency

Results of the VAS task showed that although listeners overall used the entire line when making their responses, listeners differed considerably in how they performed the VAS task, consistent with previous studies (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016). As histograms in Figure 3.3 show, some listeners made more categorical response patterns using the two endpoints of the line (Figure 3.3A1) while others made more gradient response patterns providing responses across the entire line (Figure 3.3B1). Figure 3.3 also demonstrates listeners' responses by plotting VAS responses as a function of vowel spectral quality (3.3A2 & 3.3B2) and duration (3.3A3 & 3.3B3). Specifically, Figure 3.3A shows that

[2] One participant had a very high Go/No-go score (i.e., 7.67), which is more than 3 standard deviations (i.e., 4.5) away from the Go/No-go score mean. When correlation analyses were conducted after removing this participant, only the two working memory measures were significantly correlated ($r(51) = 0.39$, $p = 0.003$), among the four cognitive ability measures in the current study.

the responses were variable at the category boundary as a function of spectral quality (3.3A2) while response patterns were not very consistent as a function of duration (3.3A3). In contrast, Figure 3.3B shows that the responses were overall more consistent with the stimuli and showed less variability in the responses as a function of spectral quality (3.3B2) although the response patterns were quite variable as a function of duration (3.3B3), if not as variable as those for Listener 255 (3.3A3). Together, the results from the VAS task show that there were substantial individual differences in speech perception gradiency. These individual differences were quantified by fitting individuals' VAS ratings using the rotated logistic function in Kapnoula et al. (2017)[3] and were subsequently used in the following section to explore whether individual differences in gradiency in speech perception correlate with individual differences in changes in cue weights from the no load to the cognitive load condition.

[3] To obtain an estimate of the slope in two dimensions, the rotated logistic function assumes a diagonal boundary line in two-dimensional space defined by the two cues. A logistic curve is then fit orthogonal to this boundary and the estimated slope of this curve is used as a measure of gradiency, in which the steeper slope the more categorical responses.

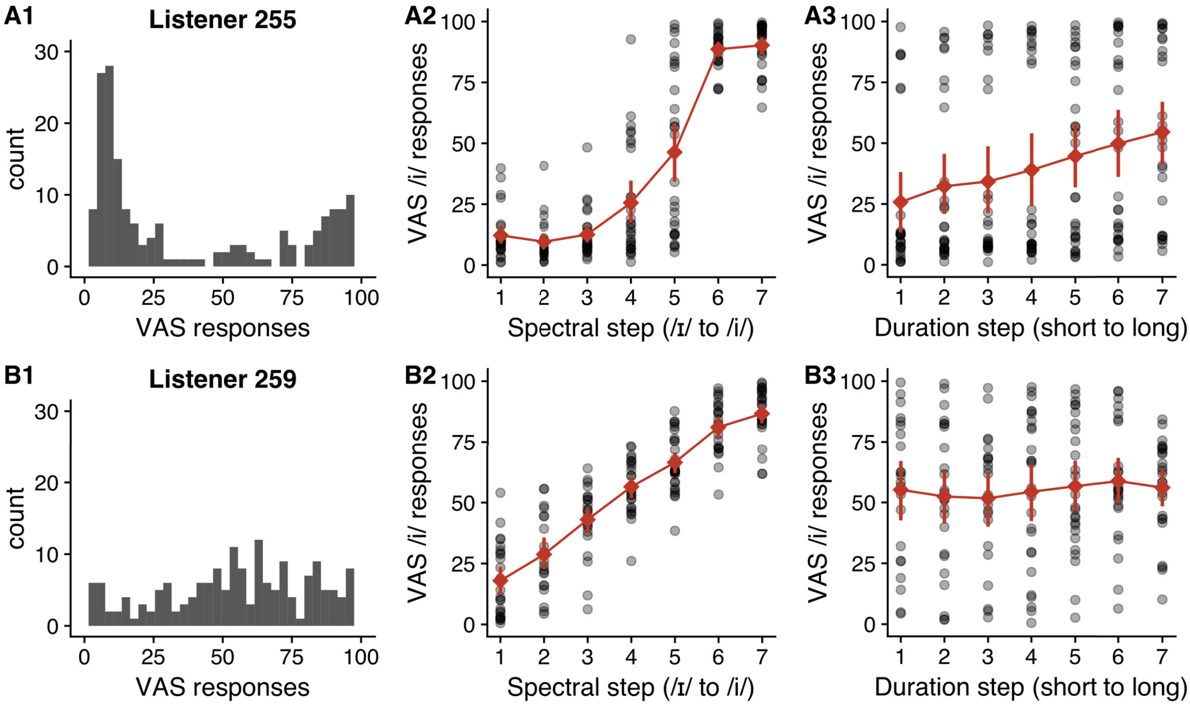

Figure 3.3. Visual analog scaling (VAS) responses by two sample listeners, Listener 255 (more categorical) and Listener 259 (more gradient).

3.3.2 The effect of cognitive load on speech perception in relation to individual difference measures

The accuracy rate of participants' visual search performance was overall well above chance (mean = 79%, standard deviation = 10%, range = 53%–95%) and similar to accuracy scores reported in Mattys and Wiget (2011), which was around 80%. To probe how performance on the visual search task affected performance on the speech perception task, correlation analyses were conducted between the magnitude of cue weight changes under cognitive load and the accuracy of the visual task. Results of the Pearson correlation showed that better performance on the visual task was correlated with an increase in spectral cue weights on the speech task ($r(52) = 0.31$, $p < 0.05$), i.e. an increase in performance on the speech task. This indicates that participants did not carry out the dual task focusing solely on one task or the other as there was no trade-off between the auditory and visual task performance.

Figure 3.4 shows the overall contribution of spectral and duration cues for categorization of /ɛ/-/æ/ across conditions. The heatmaps (top panels) show that overall listeners primarily use spectral differences to categorize the vowel contrast as indicated by changes in shading from top to bottom although there was also an effect of vowel duration in their categorization responses, as expected (Hillenbrand, Clark, & Houde, 2000; Kondaurova & Francis, 2008; 2010; Liu & Holt, 2015). Listeners' changes in cue weighting strategies across conditions are not very obvious in the heatmap representations in 3.4A, but it seems that listeners slightly increased their use of duration and that they also modified their use of spectral cues under the cognitive load condition, which is better illustrated in Figure 3.4B. The individual logistic curves in Figure 3.4B (thin lines) also suggest that there may be large differences across individuals in the extent to which listeners change their cue weighting strategies for both spectral and duration cues in the cognitive load condition.

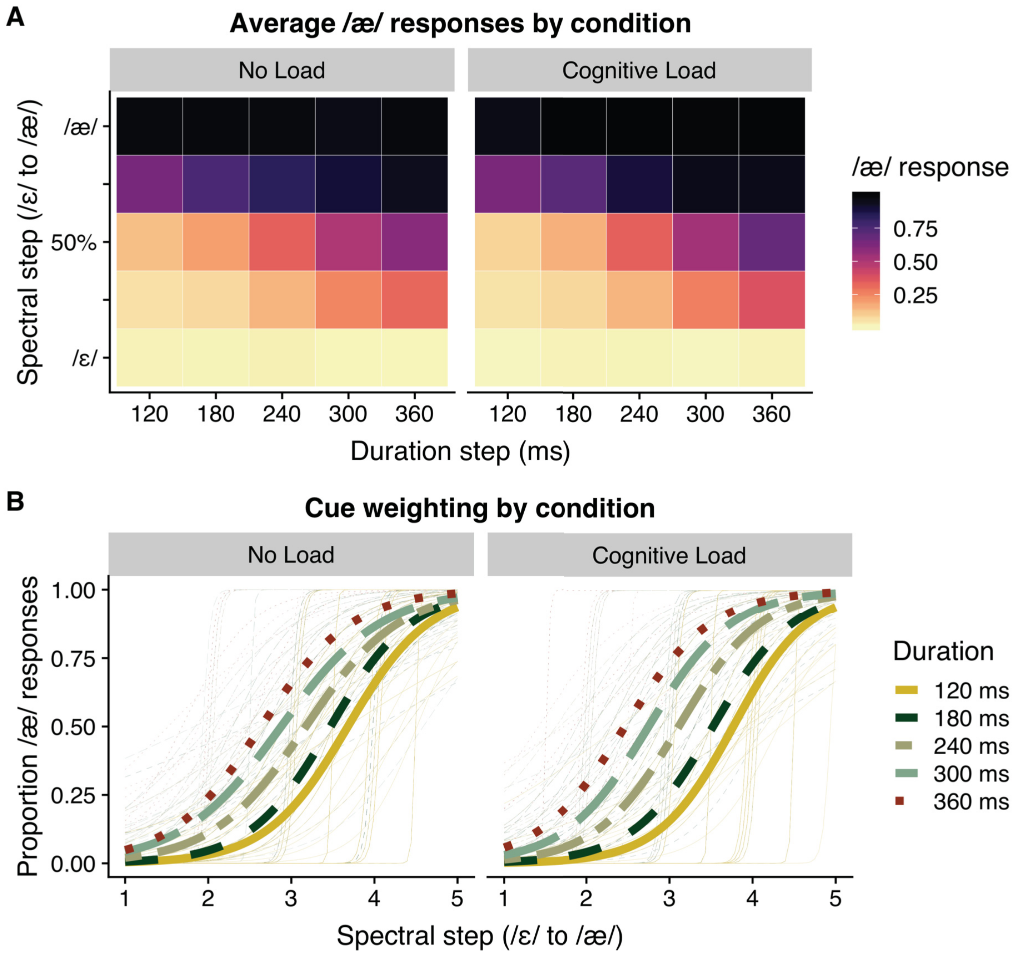

Figure 3.4. (A) Heatmap plots of overall responses for spectral quality and duration under no load and cognitive load conditions. (B) Proportion of /æ/ responses as a function of vowel spectral quality and duration differences for the no load and cognitive load conditions. Thin lines are logistic curves fit to each individual listener's data.

As presented in Table 3.2, the mixed-effects logistic regression model indicated that both vowel spectral quality ($\beta = 6.900$, $z = 30.736$, $p < 0.001$) and vowel duration ($\beta = 2.080$, $z = 15.872$, $p < 0.001$) significantly contributed to listeners' vowel categorization responses although spectral quality contributed more to vowel categorization than did vowel duration. The model also indicates that use of the cues changed across load conditions. That is, listeners' cue weighting strategies changed for both spectral ($\beta = 1.271$, $z = 4.377$, $p < 0.001$) and duration cues ($\beta = 0.831$, $z = 5.375$, $p < 0.001$) under cognitive load. Notably, the positive coefficients of the interactions between each of the cues and Condition indicate that as a group listeners increased their reliance on both cues in the cognitive load condition although there may be considerable individual differences in the extent to which individual listeners changed their cue weighting strategies under cognitive load.

Figure 3.5 illustrates the categorization responses of three listeners who demonstrated different cue weighting strategies in the cognitive load condition. The mean (SD) of difference scores of cue weights by condition is 1.21 (1.42) for the spectral cue and 0.81 (0.69) for the duration cue. Listener 261 showed an increased reliance on spectral quality in the cognitive load condition, as indicated by the steeper regression curves. In contrast, Listeners 262 showed an increased reliance on duration in the cognitive load condition, as indicated by more separation between duration curves. A third pattern of change in cue weighting is found in Listener 267 whose spectral reliance decreased in the cognitive load condition, as indicated by the overall shallower logistic curves. In the following sections, these individual differences will be examined in relation to their cognitive abilities and speech perception gradiency.

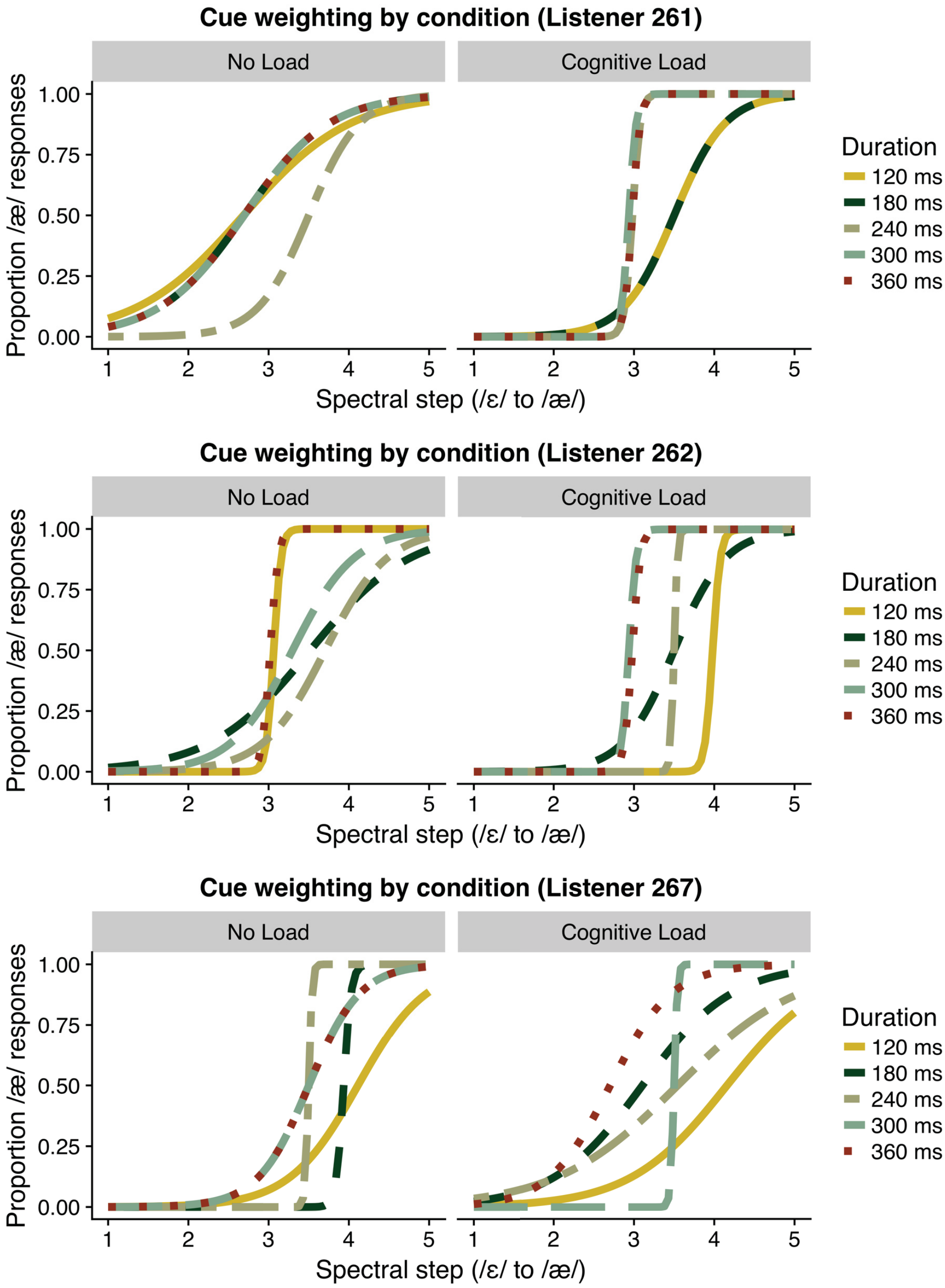

Figure 3.5. Proportion of /æ/ responses by three listeners along vowel spectral quality continuum as a function of vowel duration differences for the no load and cognitive load conditions.

Table 3.2. Summary of cognitive load model. Model coefficient estimates (β), standard errors (SE), corresponding z-values, and p-values.

Predictor	Estimate (β)	SE	z	p
Intercept	–0.472	0.139	–3.382	< 0.001
SPECTRUM	6.900	0.224	30.736	< 0.001
DURATION	2.080	0.131	15.872	< 0.001
CONDITION (No load vs. Load)	–0.019	0.108	–0.180	0.857
STROOP	–0.110	0.285	–0.387	0.698
GO/NO-GO	0.286	0.302	0.947	0.343
WM	–0.091	0.311	–0.294	0.768
VAS	–0.071	0.282	–0.253	0.800
SPECTRUM × CONDITION	1.271	0.290	4.377	< 0.001
DURATION × CONDITION	0.831	0.154	5.375	< 0.001
SPECTRUM × STROOP	–0.807	0.433	–1.861	0.062
SPECTRUM × GO/NO-GO	–1.740	0.446	–3.895	< 0.001
SPECTRUM × WM	0.294	0.479	–0.614	0.539
SPECTRUM × VAS	1.101	0.489	2.251	0.024
DURATION × STROOP	–0.396	0.262	–1.509	0.131
DURATION × GO/NO-GO	–0.323	0.271	–1.190	0.233
DURATION × WM	–0.256	0.285	–0.896	0.370
DURATION × VAS	0.511	0.285	1.793	0.073
CONDITION × STROOP	–0.302	0.208	–1.449	0.147
CONDITION × GO/NO-GO	0.173	0.218	0.795	0.426
CONDITION × WM	–0.168	0.229	–0.735	0.462
CONDITION × VAS	0.058	0.216	0.271	0.786
SPECTRUM × CONDITION × STROOP	–0.442	0.473	–0.936	0.349
SPECTRUM × CONDITION × GO/NO-GO	–1.007	0.457	–2.202	0.027
SPECTRUM × CONDITION × WM	0.139	0.520	0.269	0.787
SPECTRUM × CONDITION × VAS	0.355	0.664	0.535	0.592
DURATION × CONDITION × STROOP	–0.400	0.285	–1.402	0.160
DURATION × CONDITION × GO/NO-GO	0.141	0.274	0.516	0.605
DURATION × CONDITION × WM	0.623	0.309	2.015	0.043
DURATION × CONDITION × VAS	0.462	0.368	1.254	0.209

To investigate the effect of cognitive load on acoustic dimensions and how it relates to individual difference measures (i.e., cognitive abilities and phoneme categorization gradiency), of particular interest are three-way interactions involving acoustic dimensions (SPECTRUM and DURATION), CONDITION and individual differences measures (i.e., STROOP, GO/NO-GO, WM, and VAS). As shown in Figure 3.6A[4], there was a significant three-way interaction of SPECTRUM × CONDITION × GO/NO-GO (β = –1.007, z = –2.202, p = 0.027),

[4] Note that when a correlation analysis was conducted after removing the participant with very high Go/No-go scores (i.e., Participant 254), it did not differ from the result reported here.

indicating that individuals with better inhibitory control (as measured by the Go/No-go task)––faster response times for Go stimuli and smaller errors for No-go stimuli—also showed more spectral changes under cognitive load. The model also found a significant three-way interaction of DURATION × CONDITION × WM ($\beta = 0.623$, $z = 2.015$, $p = 0.043$) as in shown in Figure 3.6B. This indicates that individuals with better working memory also showed more duration changes in the distracting condition.

It should be noted that the model found no significant interactions of the VAS slopes and any of the changes in cue weighting strategies under cognitive load. This suggests that some listeners are more sensitive to fine-grained acoustic details such as within-category differences in speech sound categorization as previously illustrated but these differences are not linked to the magnitude of cue weights listeners adjust to compensate for the distracting condition.

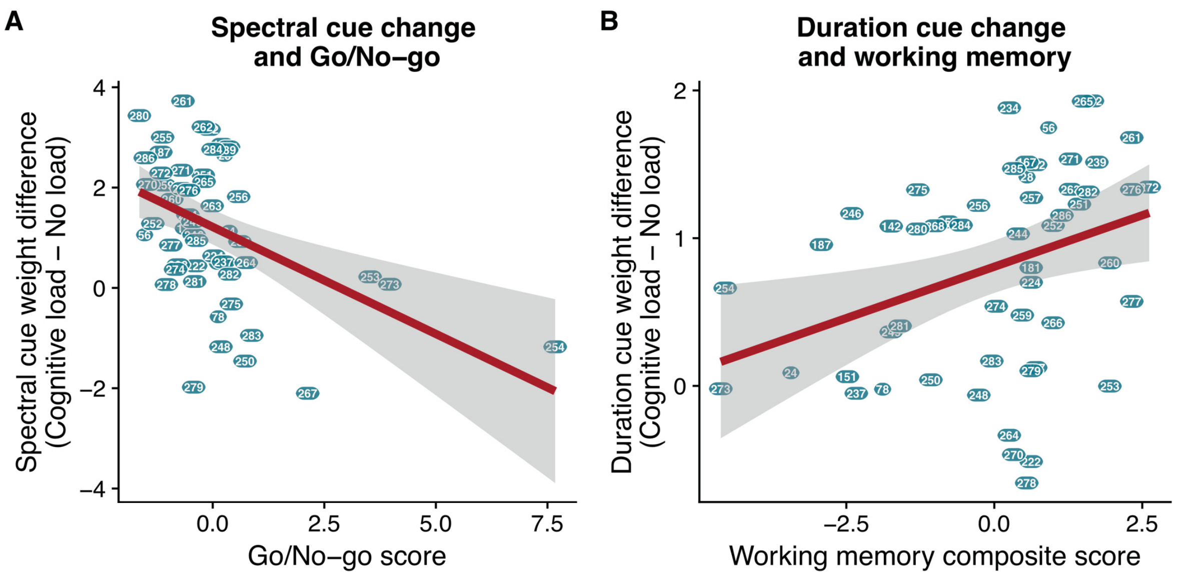

Figure 3.6. Scatter plots of individual listeners' spectral cue weight differences between cognitive load and no load conditions as a function of Go/No-go scores (A) and duration cue weight differences between cognitive load and no load conditions as a function of working memory composite scores (B).

3.4 Discussion

The aim of the current study was to investigate how and to what extent speech perception abilities are modulated by increased cognitive load in a dual task. We expected that listeners' use of multiple acoustic cues would be modulated under load conditions, but we also expected that individuals would differ in how they adapt to the distracting condition. In terms of potential sources of these individual differences, we expected that individual differences in cognitive abilities (i.e., working memory and inhibitory control) and speech perception gradiency would correlate with the patterns of cue weighting strategies under cognitive load.

3.4.1 The effect of cognitive load on speech perception

We found that listeners' cue weighting strategies were modulated under cognitive load imposed by a concurrent visual search task. Unlike findings of previous research in which cue weights given to a primary acoustic cue decreased and those given to a secondary cue increased under cognitive load (Gordon et al., 1993), the present results revealed that listeners overall showed an increased reliance on both the primary (spectral quality) and the secondary cue (vowel duration) at the group level. This may be surprising because it is plausible to assume that listeners would have more difficulty in listening when their attentional resources are limited and thus the increased cognitive load would disrupt speech perception acuity (e.g., Gordon et al., 1993; Mattys & Palmer, 2015; Mitterer & Mattys, 2017).

However, this view to some extent depends on the assumption that speech perception is a passive cognitive process. According to the noisy encoding view of speech perception under cognitive load (Gordon et al., 1993), the encoding of acoustic details would be disrupted and thus listeners would have less robust representation of acoustic cues in difficult listening conditions such as under divided attention. The noisy encoding approach to the processing of speech inputs can be interpreted as a passive system of speech perception in which acoustic-phonetic patterns are mapped onto phonological categories and these linguistic categories are relatively static and tend not to be adjusted by contextual demands (Heald & Nusbaum, 2014). In fact, recent research has failed to find that cognitive load has a significant influence on the details of acoustic-phonetic representations (Bosker et al., 2017). For example, Bosker et al. (2017) investigated whether cognitive load during processing of a precursor sentence influenced responses to a target word through context effects. They found that the spectral and temporal properties of a sentence context were not modulated under increased cognitive load although listeners' overall responses were biased towards longer target words. Recall also that Mitterer and Mattys (2017) reported no effect of cognitive load on categorization functions in a two alternative forced choice task.

Alternatively, speech perception, especially early stages of speech encoding, can also be considered an active cognitive process (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986). According to this view, changing context-specific demands on speech perception are important components of perceptual processing that contribute to the dynamics of the early encoding of speech. Thus, even early-stage speech perception processes are subject to attentional control and active cognitive processes rather being passive as some previous studies have suggested. This approach assumes the use of cognitive resources such as working memory and attention in the process of adjustments of speech category structures to adapt to the demands of contextual variability. The present results in which listeners showed increased cue weights under load—potentially indicating increased attention to the signal—may be interpreted in terms of this active cognitive processing model of speech perception.

This active perception process that can be influenced by contextual and/or attentional factors has been shown in recent work. In her study of coarticulatory vowel nasality, Zellou (2017) found that listeners show more specific encoding of coarticulatory vowel nasality in delayed repetition priming under more challenging conditions (i.e., multiple takers) than under less challenging conditions (i.e., a single talker). This may indicate that the encoding of acoustic-phonetic details can be enhanced by actively adapting to contextual demands. Similarly, Strauss and Francis (2017) found that listeners adjusted cue weighting strategies under attentional load to maximize speech sound categorization and overcome resource limitations. Overall, the findings of the current study point to adaptive plasticity of speech perception in adverse conditions. That is, listeners dynamically adjust their cue weighting strategies in response to changing contextual demands.

Crucially, this active control of speech perception relies on individual listeners' cognitive resources and varies widely across individual listeners. Indeed, at the individual level we found that there were marked differences across individuals in the effect of attentional modulation on perceptual cue weighting. That is, some listeners' cue weights were overall reduced by attentional demands while others were not. Most importantly, quite a number of individuals showed cue enhancement patterns in the cognitive load condition, giving more weights to both spectral and duration cues. This suggests that those listeners with greater enhancement of cue weights may have been able to draw on cognitive resources to minimize the influence of cognitive load on their performance while others could not. These considerable individual differences in the effect of cognitive load on speech perception may indicate that results in previous work (e.g., Bosker et al., 2017) in which there was no significant effects of cognitive load on use of multiple acoustic cues may be due in part to obscured differences across individuals in their cue weight shifts under cognitive load.

The active cognitive control view of speech perception predicts that listeners recruit cognitive resources depending on contextual constraints to support speech perception and these cognitive functions include working memory, attention, and perceptual learning mechanisms. This may indicate that individuals with better cognitive abilities adapt to challenging listening conditions to a greater extent than those with poor cognitive abilities, which may in turn contribute to their adaptive cue weighting strategies in the distracting condition.

3.4.2 The role of cognitive abilities in speech perception under cognitive load

Indeed, the results of the present study showed that listeners manifested active speech perception processes in the difficult listening condition and these active perceptual processes were supported by listeners' cognitive abilities. In order words, individual differences in compensatory changes in cue weighting strategies under cognitive load are linked to domain-general cognitive factors such as working memory and inhibitory control.

First, we found that working memory capacity is linked to changes in cue weighting strategies under cognitive load. That is, individuals with higher working memory capacity showed more adjustments of their cue weights to adapt to attentional demands than did those with lower working memory capacity. This is in line with previous research in which greater working memory capacity correlates with the perception of unfamiliar accents (Banks et al., 2015; Janse & Adank, 2012). This is also consistent with results from Francis (2010) in which individuals with increased working memory demand performed better in recognizing speech in the presence of a competing talker.

For inhibitory control, we found that listeners' ability to adapt to cognitive demands varied as a function of their inhibitory control. Individuals with greater inhibition showed more compensatory changes in utilizing multiple acoustic cues under cognitive load. This indicates that inhibitory control may facilitate selective attention and subsequent use of acoustic cues under cognitively demanding conditions. Inhibitory control has mostly been associated with speech perception processes in background noise (Tamati et al., 2013), in older adults (Janse & Adank, 2012), or in adaptation to unfamiliar speech (Banks et al., 2015). The present result provides some evidence that inhibition ability is linked to adaptive changes in cue weighting strategies in distracting conditions although more research is necessary to better understand the link between inhibitory control and speech perception under cognitive load.

Interestingly, the present result showed that inhibitory control as measured by the Go/No-go task, not by the Stroop task, was linked to individual differences in adjustments of cue weights under cognitive load. This may indicate that different inhibition tasks may tap into different components of inhibitory control. With regard to inhibition as a cognitive construct, recent studies have suggested that inhibition can be divided into its subcomponents such as the ability to suppress a dominant response and the ability to ignore distracting information (Friedman & Miyake, 2004; Howard, Johnson, & Pascual-Leone, 2014). There have also been some debates on specific properties of inhibition as a cognitive construct (Friedman & Miyake, 2017). Together, individuals with better inhibition ability may be better able to compensate for the impact of cognitive load on speech perception by utilizing cue weighting strategies, but subcomponents of inhibition ability and how they differentially influence speech perception under cognitive load require further investigation.

Furthermore, it should be noted that working memory and inhibitory control showed different patterns of links to individual differences in adjustments of cue weighting strategies under cognitive load. That is, working memory capacity was linked to a duration cue adjustment whereas inhibitory control was linked to a spectral cue adjustment. Previous work on category learning has indicated that working memory is more closely associated with rule-based category learning whereas inhibitory control is relevant to both rule-based and information-integration category learning (Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010). Although speculative, in the present study phonetic adjustment by giving more weight to the duration cue (duration cue enhancement) could be verbally defined (e.g., longer vs. shorter), which may be more closely related to verbalizable rule-based categorization. On the other hand, phonetic adjustment by giving more weight to the spectral cue (spectral cue enhancement) may be less likely to be verbalizable and the combination of different formant frequency information (e.g., F1 and F2) may be somewhat associated with information-integration categorization. Thus, these different category learning systems in their relation to general cognitive abilities may account for adjustments of cue weighting strategies and their relations to potential differences between working memory and inhibitory control.

3.4.3 The role of categorization gradiency in speech perception under cognitive load

In line with previous studies (Kapnoula et al., 2017; Kong & Edwards, 2011; 2016), the present study showed that individual listeners differed considerably in how gradiently they perceive speech sounds, but there was no evidence that these individual differences in speech perception gradiency are linked to adjustment of cue weighting strategies under divided attention. This may indicate that greater sensitivity to within-category acoustic details in speech perception may not be associated with how listeners cope with cognitive load. One explanation for this lack of relationship may be that in adverse listening conditions such as divided attention listeners are not able to use detailed acoustic information to identify sounds or segment speech and they are more likely to use higher-level lexical information to interpret the input signal (Mattys et al., 2014; Mattys & Wiget, 2011; Mattys, Brooks, & Cooke, 2009).

There has been no conclusive evidence of the role that categorization gradiency plays and whether gradient phoneme categorization is beneficial in speech perception. For example, one recent study pointed to a positive role of gradiency in speech perception (Kapnoula, Edwards, & McMurray, 2015). In an eye-tracking task using the visual world paradigm, they showed that listeners with a higher degree of categorization gradiency are better able to recover from initial misinterpretations of the speech input (i.e., lexical garden paths in their study). A handful of studies have also suggested some evidence of how categorization gradiency may be linked to other speech perception abilities and general cognitive abilities (Kapnoula et al., 2017; Kong & Edwards, 2016). Findings in these studies have suggested that individual differences in categorization gradiency are related to variability in how much listeners utilize a secondary acoustic cue (e.g., f0 in the English stop voicing contrast) and these individual differences may also be associated with differences in cognitive abilities as measured by working memory. However, post-hoc analyses of the present study showed that gradiency is not correlated with any of these measures (i.e., gradiency and baseline secondary (duration) cue use: $r(52) = 0.21$, $p = 0.14$, gradiency and working memory: $r(52) = 0.07$, $p = 0.63$, gradiency and Stroop: $r(52) = -0.13$, $p = 0.37$, gradiency and Go/No-go: $r(52) = -0.09$, $p = 0.51$). Together, these findings indicate that the functional role of speech perception gradiency remains to be fully understood.

3.5 Conclusion

The findings of the present work demonstrated that listeners' cue weighting strategies were overall modulated by increased cognitive load. More specifically, listeners revealed flexible adaptation strategies for phonetic categories in the face of cognitive load, overall increasing their reliance on both primary and secondary acoustic cues at the group level. However, there were marked differences across individuals in the extent to which these patterns manifest, and these differences were associated with individual differences in cognitive abilities (i.e., working memory and inhibitory control). The present findings are consistent with the view that speech perception is an active cognitive process (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986) and supports an emerging body of research that highlights the important role of cognitive processes in speech communication (Arlinger et al., 2009). Together, the current study provides insights into the interplay between speech and cognitive processes, leading to a more comprehensive understanding of the perceptual and cognitive mechanisms underlying flexibility in speech perception.

Chapter 4

General discussion and conclusion

4.1 Summary of the dissertation

In the present dissertation, I have presented results from two related studies that investigated individual differences in plasticity in speech perception under adverse listening conditions. The main findings of these studies are summarized below.

The goal of Study 1 was to examine whether listeners flexibly adapt to unfamiliar speech patterns such as those encountered in foreign-accented English vowels, in which the relative informativeness of acoustic dimensions (spectral quality vs. duration) is changed such that the most informative dimension (spectral quality) is no longer useful, but the role of the secondary cue (duration) is enhanced. More specifically, this study was particularly concerned with listeners' adaptive strategies to changes in the relative informativeness of acoustic dimensions such that the most informative dimension becomes no longer useful. This study further investigated whether and to what extent individual differences in cognitive abilities and phoneme categorization gradiency are related to adaptation to these atypical speech patterns. Results showed that listeners mostly used spectral quality to signal vowel category at baseline, but rapidly adapted by up-weighting reliance on duration when spectral quality became no longer informative. The VAS task showed substantial individual differences in categorization gradiency with more gradient listeners using a secondary cue more, but gradiency was not linked to degree of adaptation. Finally, results of cognitive ability tasks revealed that individual differences in inhibitory control, but not the other cognitive abilities, correlated with the amount of perceptual adaptation.

The goal of Study 2 was to better understand the mechanisms underlying plasticity in speech perception when listeners cope with a common adverse condition in everyday life (i.e., multi-tasking). More specifically, Study 2 examined how and to what extent speech perception abilities are modulated by increased cognitive load in a dual task, and whether individuals differ in the extent to which they adjust their cue weighting strategies in the utilization of multiple acoustic cues in this challenging condition. It also investigated how these individual differences in the adjustments of cue weighting strategies under cognitive load relate to individual differences in cognitive abilities (i.e., working memory and inhibition) and sensitivity to within-category differences as measured by the VAS task. Results revealed that listeners overall showed increased cue weights, which may be interpreted as a compensatory cue weighting strategy to adapt to cognitive load. However, there were large individual differences in the extent to which these compensatory cue weighting strategies manifest. The results indicated that these individual differences in compensatory strategies under cognitive load are associated with individual listeners' cognitive abilities (i.e., working memory and inhibitory control). That is, individuals with better working memory and inhibition showed greater compensatory increases in cue weights than those with poorer working memory and inhibition. These findings suggest that individuals use different compensatory strategies for utilizing multiple acoustic cues to cope with limited attentional resources, and these individual differences are linked to their cognitive abilities.

Taken together, the findings of the present dissertation add to a growing body of research addressing how individual cognitive and perceptual abilities influence speech perception. The present work confirms previous studies that there are considerable individual differences in speech perception (Idemaru et al., 2012; Kapnoula et al., 2017; Kong & Lee, 2017). However, the present results further indicate that close links between individual differences in speech perception and cognitive abilities emerge under adverse listening conditions. The present findings are consistent with the view that speech perception is an active cognitive process (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986) and support an emerging body of research that highlights the important role of cognitive processes in speech perception (Arlinger et al., 2009).

4.2 Implications for speech perception

The present dissertation addresses speech, perceptual, and cognitive processing issues that are relevant to a wide range of fields including phonetics, psycholinguistics, speech and hearing sciences, and cognitive psychology. More specifically, this dissertation explores speech and cognitive processes in combination involving mapping acoustic-phonetic information onto phonological categories, adaptive mechanisms in speech perception, and executive functions as they relate to speech perception. Understanding how these processes interact is of great interdisciplinary value and this work will provide the combination of insights that are potentially useful for obtaining a more comprehensive view of theoretical issues in speech perception and seeking practical help for methodological issues and applications in speech and hearing sciences.

The findings of both studies highlight adaptive plasticity in speech perception and indicate that speech perception is an active process (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986) rather than a passive automatic process (Gordon et al., 1993; Shiffrin & Schneider, 1977). The findings of the dissertation demonstrated that listeners flexibly accommodate and adapt to adverse conditions by making short-term adjustments of their cue weighting strategies. Specifically, Study 1 demonstrated that listeners are sensitive to distributional information of the speech input and flexibly adapt to unfamiliar speech sound categories utilizing a more informative acoustic dimension in the input while simultaneously maintaining stability of their long-term phonetic representations of their native language. Secondly, Study 2 showed that listeners' encoding of acoustic details were not overall passively impaired by cognitive load but some listeners even showed an increased weight to acoustic cues (i.e., enhanced speech perception acuity), actively responding to the contextual demands. These results indicate that listeners are sensitive to contextual constraints on speech perception and they flexibly accommodate the changes and demands of a speech context, indicating that even early stages of speech encoding are plastic and subject to cognitive control. The current work provides rare empirical evidence of an active cognitive view of speech perception (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986), especially in its interplay with cognitive processes at the individual level.

The current dissertation has potential implications for theoretical models of speech perception and for better understanding the cognitive mechanisms underlying speech processing from an interdisciplinary perspective. This work showed that individuals manifested significant differences in their compensatory cue weighting strategies under adverse conditions and these differences were associated with their cognitive abilities, indicating a link between perceptual adaptability and cognitive control. This may indicate that the compensatory mechanisms at work under challenging conditions are closely related to cognitive processes (Mattys et al., 2012). Previous work has indicated that speech perception performance can be explained by auditory measures such as pure tone thresholds in simple background noise while speech perception performance can be better explained by cognitive factors in complex or fluctuating background noise, in which perceiving speech is more demanding (Foo, Rudner, Rönnberg, & Lunner, 2007). Although these findings are based on performance on individuals with hearing aids, the interactions between auditory/speech and cognitive factors are also an important topic in a normal-hearing population for emerging fields of interdisciplinary research involving auditory cognitive science (Holt & Lotto, 2008) and cognitive hearing science (Arlinger et al., 2009). Understanding the interplay between speech perception and cognitive processing is crucial for better understanding how the speech perception system works and language interacts with cognition more generally.

4.3 Future directions

The present dissertation confirmed previous findings (Kapnoula et al., 2017; Kong & Edwards, 2016) that individuals differ considerably in the degree of within-category sensitivity in phoneme categorization and that gradiency in phoneme categorization is related to the use of secondary acoustic cues in speech perception (Study 1), extending it to a new contrast and new acoustic cues. However, this link between gradiency and secondary cue use was not found in Study 2 although individuals still varied widely in their categorization patterns. Further, Study 1 and Study 2 showed that categorization gradiency is not related to perceptual adaptation to an unfamiliar accent (Study 1) and speech perception under cognitive load (Study 2). These results raise questions about the functional role of gradiency in phoneme categorization. Although a recent study by Kapnoula et al. (2015) suggested that speech perception gradiency may be beneficial for recovering from initial misinterpretations of the speech input, more work is needed to better understand the utilization of continuous acoustic information and how this continuous use of acoustic information influences speech perception processes.

The picture that emerges from the results from the two studies is that some individuals are better able to adapt to adverse conditions by making use of acoustic information in the speech signal, suggesting that these individuals may have superior abilities in using bottom-up acoustic information in speech perception. Although bottom-up acoustic processing is crucial in speech perception, understanding speech inherently involves the subtle interplay between bottom-up acoustic and top-down linguistic (i.e., higher-level linguistic knowledge) processing (e.g., Bradlow & Alexander, 2007; Pichora-Fuller, 2008; Schertz & Hawthorne, 2018). For example, Pichora-Fuller (2008) showed that use of semantic context aids speech perception processes in challenging conditions by lowering cognitive load when a mismatch between the speech signal and meaning occurs. Schertz and Hawthorne (2018) examined whether higher-level contextual information interacts with lower-level acoustic information in speech perception and found that listeners increased reliance on contextual information when listening to a talker with a foreign accent. However, prior research has also indicated that there are significant differences in the extent to which individuals utilize top-down linguistic information in speech perception (Ishida, Samuel, & Arai, 2016). These previous findings indicate that perceiving speech in adverse conditions is challenging and accommodating these challenging conditions may require integration of multiple sources of information including bottom-up and top-down information, and individuals differ in the extent to which they utilize bottom-up acoustic and top-down linguistic information to understand speech. Thus, future work would benefit from investigating individual differences in relative use of top-down and bottom-up information in speech perception and how they are linked to the allocation of cognitive resources during speech perception.

Speech perception performance in adverse conditions may be compensated for, supported by listeners' cognitive abilities. However, using more cognitive processing resources for perceptual compensation may be related to more effortful listening although this effortful listening may not be obvious based on behavioral measures as in the current work. A related issue involving cognitive processing during speech perception is listening effort (e.g., Lemke & Besser, 2016; McGarrigle et al., 2014; Strauss & Francis, 2017). Listening effort refers to the attentional or cognitive resources required for successful speech comprehension (Strauss & Francis, 2017). In recent years, much attention has been paid to the concepts and the underlying mechanisms of listening effort. For example, Francis, MacPherson, Chandrasekaran, and Alvar (2016) suggested that listeners' different compensatory strategies to cope with the different sources of reduced intelligibility in degraded listening conditions can be measured by physiological responses in terms of listening effort. Future research would benefit from combining the behavioral measures with physiological or neurophysiological measures to better understand the nature of listening effort and its link to cognitive abilities in speech perception.

4.4 Conclusion

While speech perception is seemingly effortless, listeners constantly cope with a variety of adverse conditions in everyday life. Adopting an individual differences approach, the present work demonstrated how individuals accommodate and flexibly adapt to challenging listening conditions and suggested potential mechanisms underlying this remarkable achievement. The results of the present work confirm previous studies, indicating that there are considerable individual differences in the perception and perceptual adaptability of speech sound categories (Kapnoula et al., 2017; Kong & Edwards, 2016; Schertz et al., 2016). Furthermore, these individual differences were associated with individuals' cognitive abilities. These findings are consistent with the view that speech perception is an active cognitive process (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986) and support an emerging body of research that highlights the important role of cognitive processes in speech perception (Arlinger et al., 2009; Pichora-Fuller et al., 2016; Rönnberg et al., 2013). Taken together, the current dissertation provides unique insights into ways in which speech perception is actively interacted with and supported by cognitive processes, advancing our understanding of the mechanisms underlying plasticity in speech perception.

References

Adank, P. M., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*(3), 736–740.

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47 Suppl 2*, S53–S71. http://doi.org/10.1080/14992020802301142

Amer, T., Anderson, J. A. E., & Hasher, L. (2018). Do young adults show conceptual knowledge of previous distractors? *Memory*, *26*(2), 251–259. http://doi.org/10.1080/09658211.2017.1347187

Amer, T., Campbell, K. L., & Hasher, L. (2016). Cognitive control as a double-edged sword. *Trends in Cognitive Sciences*, *20*(12), 905–915. http://doi.org/10.1016/j.tics.2016.10.002

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of Cognitive Hearing Science. *Scandinavian Journal of Psychology*, *50*(5), 371–384. http://doi.org/10.1111/j.1467-9450.2009.00753.x

Azadpour, M., & Balaban, E. (2015). A proposed mechanism for rapid adaptation to spectrally distorted speech. *Journal of the Acoustical Society of America*, *138*(1), 44–57. http://doi.org/10.1121/1.4922226

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839. http://doi.org/10.1038/nrn1201

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, *133*(3), EL174–EL180. http://doi.org/10.1121/1.4789864

Banks, B., Gowen, E., Munro, K. J., & Adank, P. M. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America*, *137*(4), 2015–2024. http://doi.org/10.1121/1.4916265

Beddor, P. S. (2015). The relation between language users' perception and production repertoires. *Proceedings of the 18th International Congress of the Phonetic Sciences*, Glasgow, UK.

Bender, A. D., Filmer, H. L., Garner, K. G., Naughtin, C. K., & Dux, P. E. (2016). On the relationship between response selection and response inhibition: An individual differences approach. *Attention, Perception, & Psychophysics*, *78*(8), 2420–2432. http://doi.org/10.3758/s13414-016-1158-8

Bent, T., Baese-Berk, M. M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *Journal of the Acoustical Society of America*, *140*(5), 3775–3786. http://doi.org/10.1121/1.4966677

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise: A review. *Trends in Hearing*, *17*(2), 75–93. http://doi.org/10.1177/1084713813495459

Bezdjian, S., Baker, L. A., Lozano, D. I., & Raine, A. (2009). Assessing inattention and impulsivity in children during the Go/NoGo task. *British Journal of Developmental Psychology*, *27*(2), 365–383. http://doi.org/10.1348/026151008X314919

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program].

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176. http://doi.org/10.1016/j.jml.2016.12.002

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*(4), 2339–2349. http://doi.org/10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. http://doi.org/10.1016/j.cognition.2007.04.005

Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication*, *52*(11-12), 980–995. http://doi.org/10.1016/j.specom.2010.08.010

Cainer, K. E., James, C., & Rajan, R. (2008). Learning speech-in-noise discrimination in adult humans. *Hearing Research*, *238*(1-2), 155–164. http://doi.org/10.1016/j.heares.2007.10.001

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. http://doi.org/10.1016/j.wocn.2005.08.003

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658. http://doi.org/10.1121/1.1815131

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. http://doi.org/10.1016/j.cognition.2008.04.004

Cohen, R. A. (2014). *The Neuropsychology of Attention* (2nd ed.). Boston, MA: Springer US. http://doi.org/10.1007/978-0-387-72639-7

Conners, C. K., Epstein, J. N., Angold, A., & Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of Abnormal Child Psychology*, *31*(5), 555–562. http://doi.org/10.1023/A:1025457300409

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*(4), 741–773. http://doi.org/10.1111/lang.12161

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. http://doi.org/10.1146/annurev-psych-113011-143750

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, *4*(4), 500–503. http://doi.org/10.1037//0882-7974.4.4.500

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers*. Unpublished M. Sc. thesis, University of Edinburgh.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452–465. http://doi.org/10.1016/j.wocn.2009.07.006

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255. http://doi.org/10.1044/1092-4388(2007/087)

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. http://doi.org/10.1006/jpho.1997.0052

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, *18*(7), 618–631. http://doi.org/10.3766/jaaa.18.7.8

Fox, C. J., Mueller, S. T., Gray, H. M., Raber, J., & Piper, B. J. (2013). Evaluation of a short-form of the Berg Card Sorting Test. *PloS ONE*, *8*(5), e63885. http://doi.org/10.1371/journal.pone.0063885

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, & Psychophysics*, *72*(2), 501–516. http://doi.org/10.3758/APP.72.2.501

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349. http://doi.org/10.1037/0096-1523.28.2.349

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*(6), 1360–1374. http://doi.org/10.3758/APP.71.6.1360

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America*, *124*(2), 1234–1251. http://doi.org/10.1121/1.2945161

Francis, A. L., MacPherson, M. K., Chandrasekaran, B., & Alvar, A. M. (2016). Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology*, *7*(878). http://doi.org/10.3389/fpsyg.2016.00263

Franken, M. K., Eisner, F., Schoffelen, J.-M., Acheson, D. J., Hagoort, P., & McQueen, J. M. (2017). Audiovisual recalibration of vowel categories (pp. 655–658). *Interspeech 2017*, ISCA: ISCA. http://doi.org/10.21437/Interspeech.2017-122

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. http://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. http://doi.org/10.1016/j.cortex.2016.04.023

Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, *6*. http://doi.org/10.3389/fnagi.2014.00347

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612. http://doi.org/10.1146/annurev.psych.49.1.585

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*(1), 1–42. http://doi.org/10.1006/cogp.1993.1001

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear and Hearing*, *37*(5), 593–602. http://doi.org/10.1097/AUD.0000000000000316

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, *7*. http://doi.org/10.3389/fnsys.2013.00126

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, *52*(6), 1510–1529. http://doi.org/10.1044/1092-4388(2009/08-0220)

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics*, *49*(2), 187–200. http://doi.org/10.3758/BF03205038

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*. http://doi.org/10.3389/fnsys.2014.00035

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *103*(3), 1–21. http://doi.org/10.3758/s13428-017-0935-1

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, *108*(6), 3013–3022. http://doi.org/10.1121/1.1323463

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059–3071. http://doi.org/10.1121/1.2188377

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42–46. http://doi.org/10.1111/j.1467-8721.2008.00545.x

Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, *31*, 1–21. http://doi.org/10.1016/j.cogdev.2014.03.001

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. http://doi.org/10.1080/23273798.2015.1047459

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956. http://doi.org/10.1037/a0025641

Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1009–1021. http://doi.org/10.1037/a0035269

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America*, *132*(6), 3950–3964. http://doi.org/10.1121/1.4765076

Ishida, M., Samuel, A. G., & Arai, T. (2016). Some people are "more lexical" than others. *Cognition*, *151*, 68–75. http://doi.org/10.1016/j.cognition.2016.03.008

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology, and Cognition*, *19*(6), 741–758. http://doi.org/10.1080/13825585.2011.652590

Janse, E., & Adank, P. M. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563–1585. http://doi.org/10.1080/17470218.2012.658822

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, *67*(9), 1842–1862. http://doi.org/10.1080/17470218.2013.879391

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. http://doi.org/10.1080/17470218.2014.964736

Kapnoula, E. C., Edwards, J. R., & McMurray, B. (2015). Gradient categorization of speech sounds helps listeners recover from lexical garden paths. Poster presented at the *14th Auditory Perception, Cognition, and Action Meeting*, November 19, Chicago, USA.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J. R., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594–1611. http://doi.org/10.1037/xhp0000410

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. *Proceedings of Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference* (pp. 111–120).

Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, Poznań, Poland.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, *124*(6), 3959–3971. http://doi.org/10.1121/1.2999341

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics*, *38*(4), 569–587. http://doi.org/10.1016/j.wocn.2010.08.003

Kong, E. J., & Edwards, J. R. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong.

Kong, E. J., & Edwards, J. R. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. http://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J., & Lee, H. (2017). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*. http://doi.org/10.1177/0023830917729840

Kortte, K. B., Horner, M. D., & Windham, W. K. (2010). The Trail Making Test, Part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*(2), 106–109. http://doi.org/10.1207/S15324826AN0902_5

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178. http://doi.org/10.1016/j.cogpsych.2005.05.001

Lehet, M., & Holt, L. L. (2017). Dimension-based statistical learning affects both speech perception and production. *Cognitive Science*, *41*, 885–912. http://doi.org/10.1111/cogs.12413

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and Hearing*, *37*, 77S–84S. http://doi.org/10.1097/AUD.0000000000000304

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*, 320–331.

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, *47*, 36–46. http://doi.org/10.1016/j.wocn.2014.09.001

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 3–11. http://doi.org/10.1177/002383098602900102

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783–1798. http://doi.org/10.1037/xhp0000092

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. http://doi.org/10.1037/0033-2909.109.2.163

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia*, *48*(10), 2998–3008. http://doi.org/10.1016/j.neuropsychologia.2010.06.008

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409. http://doi.org/10.1037/0096-1523.33.2.391

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, *2*(1), 15–35. http://doi.org/10.1016/0167-6393(83)90061-4

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. http://doi.org/10.3758/s13428-011-0168-7

Mattys, S. L., & Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *Journal of the Acoustical Society of America*, *137*(3), 1464–1472. http://doi.org/10.1121/1.4913507

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160. http://doi.org/10.1016/j.jml.2011.04.004

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, *21*(3), 748–754. http://doi.org/10.3758/s13423-013-0544-7

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. http://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7/8), 953–978.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. http://doi.org/10.1037/0096-3445.134.4.477

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology*, *53*(7), 433–445. http://doi.org/10.3109/14992027.2014.890296

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. http://doi.org/10.1037/a0011747

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. http://doi.org/10.1016/j.cogpsych.2009.06.003

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. http://doi.org/10.1207/s15516709cog0000_79

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, *79*(1), 344–351. http://doi.org/10.3758/s13414-016-1195-3

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. http://doi.org/10.1177/0963721411429458

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000a). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 169–183. http://doi.org/10.1055/s-2000-7563

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000b). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250–259. http://doi.org/10.1016/j.jneumeth.2013.10.024

Munson, B., Schellinger, S. K., & Edwards, J. R. (2017). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, *31*(1), 56–79. http://doi.org/10.1080/02699206.2016.1233292

Nagaraj, N. K., & Magimairaj, B. M. (2017). Role of working memory and lexical knowledge in perceptual restoration of interrupted speech. *Journal of the Acoustical Society of America*, *142*(6), 3756–3766. http://doi.org/10.1121/1.5018429

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. http://doi.org/10.1016/S0010-0285(03)00006-9

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In *Pattern Recognition by Humans and Machines* (pp. 113–157). Academic Press. http://doi.org/10.1016/B978-0-12-631403-8.50009-6

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*(sup2), S72–S82.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S–27S. http://doi.org/10.1097/AUD.0000000000000312

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*. http://doi.org/10.3389/fnsys.2013.00031

Scharenborg, O., Weber, A., & Janse, E. (2015). Age and hearing loss and the use of acoustic cues in fricative categorization. *Journal of the Acoustical Society of America*, *138*(3), 1408–1417. http://doi.org/10.1121/1.4927728

Schellinger, S. K., Munson, B., & Edwards, J. R. (2017). Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics*, *31*(1), 80–103. http://doi.org/10.1080/02699206.2016.1205665

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3-4), 249–263. http://doi.org/10.1016/j.wocn.2013.03.007

Schertz, J., & Hawthorne, K. (2018). The effect of sentential context on phonetic categorization is modulated by talker accent and exposure. *Journal of the Acoustical Society of America*, *143*(3), EL231–EL236. http://doi.org/10.1121/1.5027512

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. http://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*(1), 355–367. http://doi.org/10.3758/s13414-015-0987-1

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology*, *5*. http://doi.org/10.3389/fpsyg.2014.01475

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. http://doi.org/10.1037//0033-295x.84.2.127

Souza, P., & Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *International Journal of Audiology*, *54*(10), 705–713. http://doi.org/10.3109/14992027.2015.1043062

Sörqvist, P., & Rönnberg, J. (2014). Individual differences in distractibility: An update and a model. *PsyCh Journal*, *3*(1), 42–57. http://doi.org/10.1002/pchj.47

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(4), 809–825. http://doi.org/10.3758/s13415-017-0513-0

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology*, *24*(7), 616–634. http://doi.org/10.3766/jaaa.24.7.10

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464. http://doi.org/10.1111/j.1551-6709.2009.01077.x

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540. http://doi.org/10.1177/0956797610384142

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654. http://doi.org/10.1080/09658210902998047

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*(1), 57–79. http://doi.org/10.1348/000712604322779460

Weeks, J. C., Biss, R. K., Murphy, K. J., & Hasher, L. (2016). Face–name learning in older adults: A benefit of hyper-binding. *Psychonomic Bulletin & Review*, *23*(5), 1559–1565. http://doi.org/10.3758/s13423-016-1003-z

Wickens, C. D., & McCarley, J. S. (2008). *Applied Attention Theory*. Boca Raton, FL: CRC Press.

Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing*, *37*, 35S–43S. http://doi.org/10.1097/AUD.0000000000000310

Yu, A. C. L., & Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *Journal of the Acoustical Society of America*, *136*(1), 382–388. http://doi.org/10.1121/1.4883380

Zellou, G. (2017). Specificity for coarticulatory vowel nasality in lexical representations. *Journal of the Acoustical Society of America*, *142*(4), EL375–EL380. http://doi.org/10.1121/1.5006817

Zhang, X., & Samuel, A. G. (2014). Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 200–217. http://doi.org/10.1037/a0033182