Journal of Phonetics 81 (2020) 100984

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Individual differences in perceptual adaptation to unfamiliar phonetic categories

Donghyun Kim [a,*], Meghan Clayards [b,c], Eun Jong Kong [d]

[a] Department of Psychology, University of Exeter, Exeter, United Kingdom
[b] Department of Linguistics, McGill University, Montreal, Canada
[c] School of Communication Sciences and Disorders, McGill University, Montreal, Canada
[d] Department of English, Korea Aerospace University, Gyeonggido, South Korea

ARTICLE INFO

Article history:
Received 14 August 2019
Received in revised form 4 May 2020
Accepted 5 May 2020

Keywords:
Speech perception
Perceptual learning
Cue weighting
Individual differences
Categorization gradiency
Cognitive abilities

ABSTRACT

The present study examines whether listeners flexibly adapt to unfamiliar speech patterns such as those encountered in foreign-accented English vowels, where the relative informativeness of primary (spectral quality) and secondary (duration) cues tends to be reversed (e.g., spectrally similar but exaggerated duration differences between *bet* and *bat*). This study further tests whether listeners' adaptive strategies are related to individual differences in phoneme categorization gradiency and cognitive abilities. Native English listeners (N = 36) listened to a continuum of vowels from /ɛ/ to /æ/ (as in *head* and *had*) varying in spectral and duration values to complete a perceptual adaptation task and a visual analog scaling (VAS) task. Participants also completed cognitive tasks examining executive function capacities. Results showed that listeners mostly used spectral quality to signal vowel category at baseline, but flexibly adapted by up-weighting reliance on duration when spectral quality became no longer diagnostic. In the VAS task, some listeners made more categorical responses while others made more gradient responses in vowel categorization, but these differences were not linked to their adaptive patterns. Results of cognitive tasks revealed that individual differences in inhibitory control correlated, to some degree, with the amount of adaptation. Together, these findings suggest that listeners flexibly adapt to unfamiliar speech categories using distributional information in the input and individual differences in cognitive abilities may influence their adaptability.

1. Introduction

When perceiving speech, listeners face an enormous amount of variability in phonetic realization. This variability may 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 may sound unfamiliar. For example, one talker's /ʃ/ as in *ship* can sound very much like another talker's /s/ as in *sip* (Newman, Clouse, & Burnham, 2001). The English word *ship* can also be pronounced like *sheep* with a vowel closer to /i/ rather than /ɪ/ by non-native speakers of English such as Spanish and Korean speakers (Flege, Bohn, & Jang, 1997). 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). The goal of the present study is to better understand this remarkably flexible process. Specifically, 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.

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 et al., 2001). Despite this variability in speech sound realization, listeners can often overcome initial difficulties and show intelligibility improvements with

* Corresponding author.
E-mail address: d.kim2@exeter.ac.uk (D. Kim).

https://doi.org/10.1016/j.wocn.2020.100984
0095-4470/

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. 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). These studies have mostly focused on phonetic categories and how they are retuned to cope with acoustic–phonetic variability. Given the considerable variability inherent in the speech signal, 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 (e.g., Kraljic & Samuel, 2005; McQueen et al., 2006; Norris et al., 2003). 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. Previous research has further 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 (e.g., 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 phonetic 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 their use of f0 in English stop voicing categorization. That is, listeners decreased their reliance on f0 when it was no longer useful in defining voicing categories. 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 (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 distributional information in the input to learn which dimension is most reliable. In particular, we hypothesize that listeners can increase their use of a secondary dimension when the most reliable dimension is no longer informative.

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 et al., 1997). A specific example would be that when native speakers of Korean pronounce English /æ/ as in *bat* and /ɛ/ as in *bet*, they are likely to make /æ/ exaggeratedly long and /ɛ/ very short, while producing the two vowels with more similar quality than a native English speaker would. This can cause intelligibility problems for native listeners of English.

In the present study, native English listeners were exposed to unfamiliar speech which sounds like foreign-accented English vowels that deviate from English norms in the informativeness of the primary acoustic dimension. That is, they were exposed to an uninformative primary acoustic dimension (spectral quality) while the secondary acoustic dimension (vowel duration) remained informative. Listeners are expected to adapt to this unfamiliar speech pattern by redirecting their attention to the most diagnostic acoustic dimension (i.e., vowel duration) when categorizing the vowels (cf. Francis & Nusbaum, 2002).

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., Clayards, 2018; 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, Coetzee, Styler, McGowan, & Boland, 2018; Idemaru et al., 2012; Kapnoula et al., 2017; Kong & Edwards, 2011, 2016). For example, Kong and Edwards (2011, 2016) examined perceptual 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 contrast. Beddor et al. (2018) 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) is one study that examined individual differences in perceptual adaptation to foreign sound categories in the use of multiple acoustic cues. They investigated whether non-native listeners show adjustments to their cue weighting strategies in response to changes in the speech input using the dimension-based learning paradigm described above. They also tested whether adaptation patterns are related to individual 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 differences in initial cue weighting strategies result in different patterns of adaptation. 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.

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 task (forward or backward), 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; Lev-Ari & Peperkamp, 2013; 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 accented speech.

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, Baese-Berk, Borrie, & McKee, 2016; Janse & Adank, 2012; Kim & Hazan, 2010; Kong & Edwards, 2016). For example, Kim and Hazan (2010) adopted several cognitive ability tasks such as inhibitory control, 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. In Bent et al. (2016), cognitive factors were examined in relation to 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 in the study including inhibitory control and cognitive flexibility. Similarly, Kong and Edwards (2016) found no significant relation between cognitive measures such as inhibitory control 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), the present study further 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 individual listeners' cognitive abilities in speech perception as to whether cognitive abilities contribute to better adaptation to unfamiliar speech patterns.

1.4. Categorization gradiency in speech perception

Although cognitive abilities may play a role in flexibility in speech perception, listeners' sensitivity to acoustic details may also contribute to better adaptation to variability in speech (Kim & Hazan, 2010). 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 are 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 individuals considerably differ in the gradiency of their perceptual judgments and importantly that gradient listeners' responses more closely reflect subtle acoustic differences in the stimuli. In other words, more gradient listeners' responses shifted from one end of the scale to the other as the stimuli continuously varied along multiple acoustic dimensions.

Another 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 by extending them to the perception of vowels. Vowel perception has sometimes been described as more gradient than consonants (Schouten, Gerrits, & van Hessen, 2003), so the same patterns of individual variability may not hold. Furthermore, this study investigates whether gradiency predicts listeners' patterns of perceptual adaptation.

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 realizations of speech sounds 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 (i.e., an ambiguous primary cue and an enhanced secondary cue). 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 phonetic categories. Research questions of this study are:

1. Do listeners flexibly adapt to unfamiliar speech that deviates from learned long-term regularities by increasing their reliance on a secondary acoustic dimension when the most informative dimension is no longer diagnostic?
2. Are previously observed patterns of individual differences (i.e., more gradient vs. more categorical) in consonant categorization also observed for vowel categorization? If so, are patterns of cate-

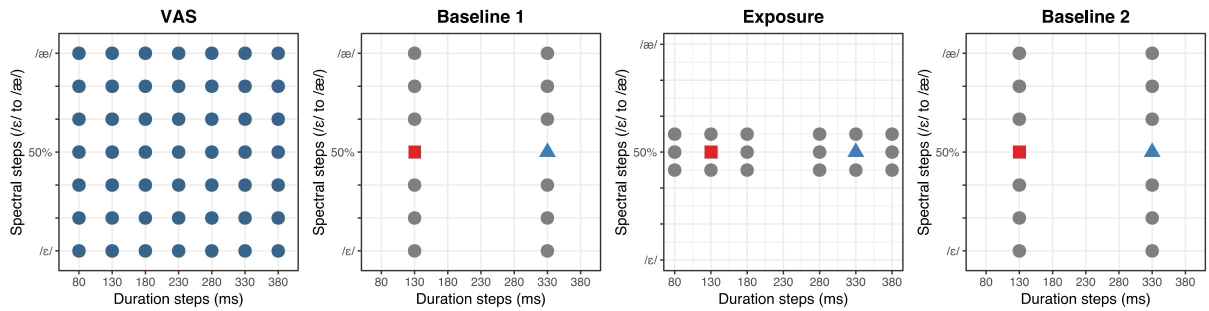

Fig. 1. Illustration of the stimuli used in the VAS and the adaptation task. Test stimuli for the adaptation task were the red square and the blue triangle. Baseline was repeated after Exposure (as in Baseline 1 and Baseline 2).

gorization 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 up-weight 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 use a secondary dimension to 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 may play a role in their adaptive patterns. Although there was some evidence of the link between cognitive abilities and speech perception, previous work has reported no strong relationship or inconsistent findings (e.g., Bent et al., 2016; Kim & Hazan, 2010; Kong & Edwards, 2016). Thus, we do not have strong specific hypotheses involving cognitive abilities. Broadly, we expect that better cognitive abilities, particularly better working memory capacity and inhibitory control, may help listeners adapt to unfamiliar phonetic categories, but the analyses of cognitive abilities will be quite exploratory in nature. 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 phonetic categories.

2. Methods

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. Stimuli

Fig. 1 illustrates stimuli for the VAS task, and Baseline 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 (approximately 50% *had* responses) was chosen. Based on the two end-point tokens and the most ambiguous step, auditorily and acoustically distinct intermediate steps (e.g., at least two spectral steps apart) out of the twenty-step continuum were selected to make the seven-step spectral continuum.[1] 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 $\times$ 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 stimuli for the adaptation task 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 Fig. 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 duration across the course of the experiment.

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

[1] All stimuli and additional information are available on the Open Science Framework at https://osf.io/5mfea/.

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.

2.3.1. 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 min.

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, Emerson, & Friedman, 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 min. The cognitive tasks in total took approximately 30 min.

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. During the task, participants did not receive any feedback on their performance. All the trials within a block were randomly presented through headphones at a comfortable listening level. The adaptation task was completed in approximately 25 min.

2.4. Analysis

2.4.1. Vowel categorization at Baseline 1

As in previous work (Kapnoula et al., 2017), the relationship between secondary cue use and categorization gradiency was investigated using differences in crossover point for the two continua (short and long vowel) from the baseline 2AFC task, as illustrated in Fig. 2. Crossover points were measured for

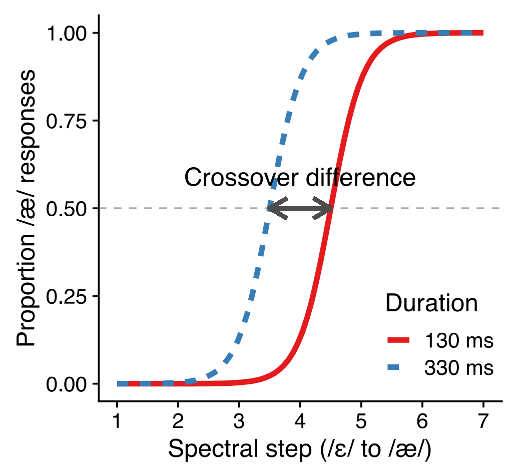

Fig. 2. Hypothetical illustration of duration cue use at Baseline 1 as measured by the difference in 2AFC crossover points between short (130 ms) and long (330 ms) vowel durations.

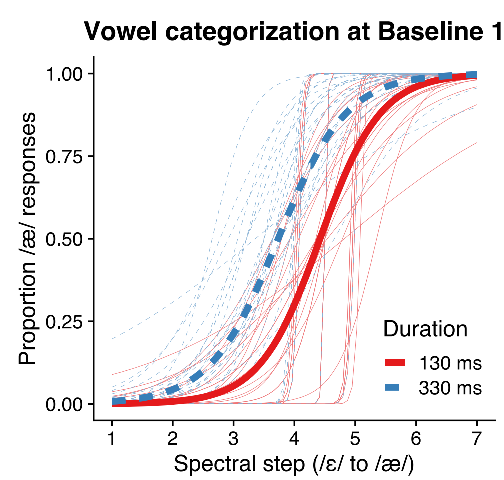

Fig. 3. Proportion of /æ/ responses along vowel spectral quality continuum at Baseline 1 as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are logistic curves fit to each individual listener data for each vowel duration.

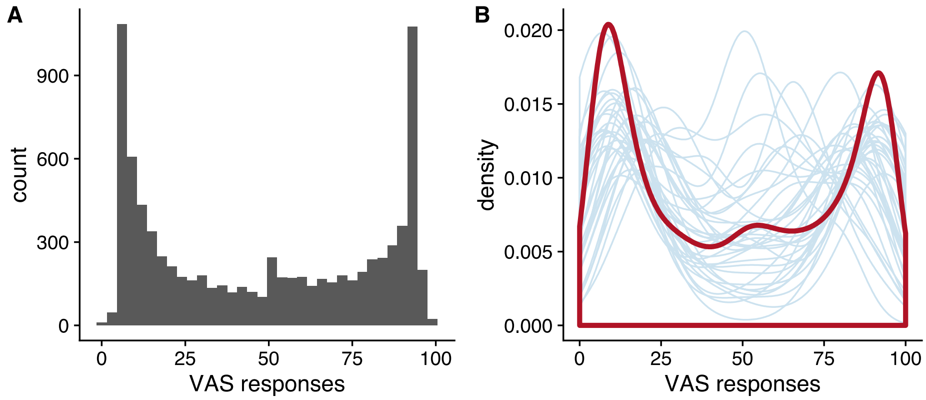

Fig. 4. (A) A histogram of overall visual analog scaling (VAS) responses. (B) Density curves showing the distributions of the overall average (thick red) and individual (thin blue) VAS responses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each participant by fitting a four-parameter (i.e., minimum and maximum asymptotes, slope, and crossover) logistic function and using the midpoint variable (see Kapnoula et al., 2017 for details). For the purpose of the present study, 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, 2010).

2.4.2. 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. To quantify degree of gradiency for each individual, a rotated logistic function was fit following Kapnoula et al. (2017).[2] Gradiency was assessed using the slope of the rotated logistic function (shallower slopes—smaller values—indicate more gradient responses).

2.4.3. 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 less 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, the cognitive measures will be entered as predictors in a multiple linear regression analysis to assess how they are associated with the gradiency of response on the VAS task.

2.4.4. 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 (Jaeger, 2008) will be used to investigate whether individual listeners' responses to Test stimuli are predicted by individual difference measures of phoneme categorization gradiency, duration cue use and cognitive ability measures (inhibitory control, working memory, cognitive flexibility, and sustained attention), using the *glmer()* function from the *lme4* package (ver.1.1-16) in R (Bates, Mächler, Bolker, & Walker, 2015; R Core and Team, 2017). Statistical models will be described more in detail in the results section.

3. Results

3.1. Vowel categorization at Baseline 1

Fig. 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 (Hillenbrand et al., 2000; Kondaurova & Francis, 2008, 2010; Liu & Holt, 2015). To assess this pattern of vowel categorization at baseline, we performed a separate mixed-effects logistic regression analysis with random intercepts and random slopes for spectral and duration steps for participants. Spectral steps were standardized by centering and dividing by two standard deviations and duration steps were centered by subtracting the mean in the model (Gelman, 2008). This analysis confirmed that listeners primarily rely on vowel spectral quality (β = 10.167, *SE* = 0.740, *z* = 13.721, $p < 0.001$) although vowel

[2] 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.

duration also contributes to vowel categorization (β = 0.405, *SE* = 0.054, *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 Fig. 3 also indicates considerable individual differences in the use of vowel duration for vowel categorization, which will be discussed in relation to gradiency in phoneme categorization in Section 3.2.3.

3.2. Gradiency and cognitive measures

3.2.1. The VAS task

Fig. 4A 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. Fig. 4B illustrates the distributions of VAS responses for each participant, with the overall average distribution superimposed. The individual density curves show con-

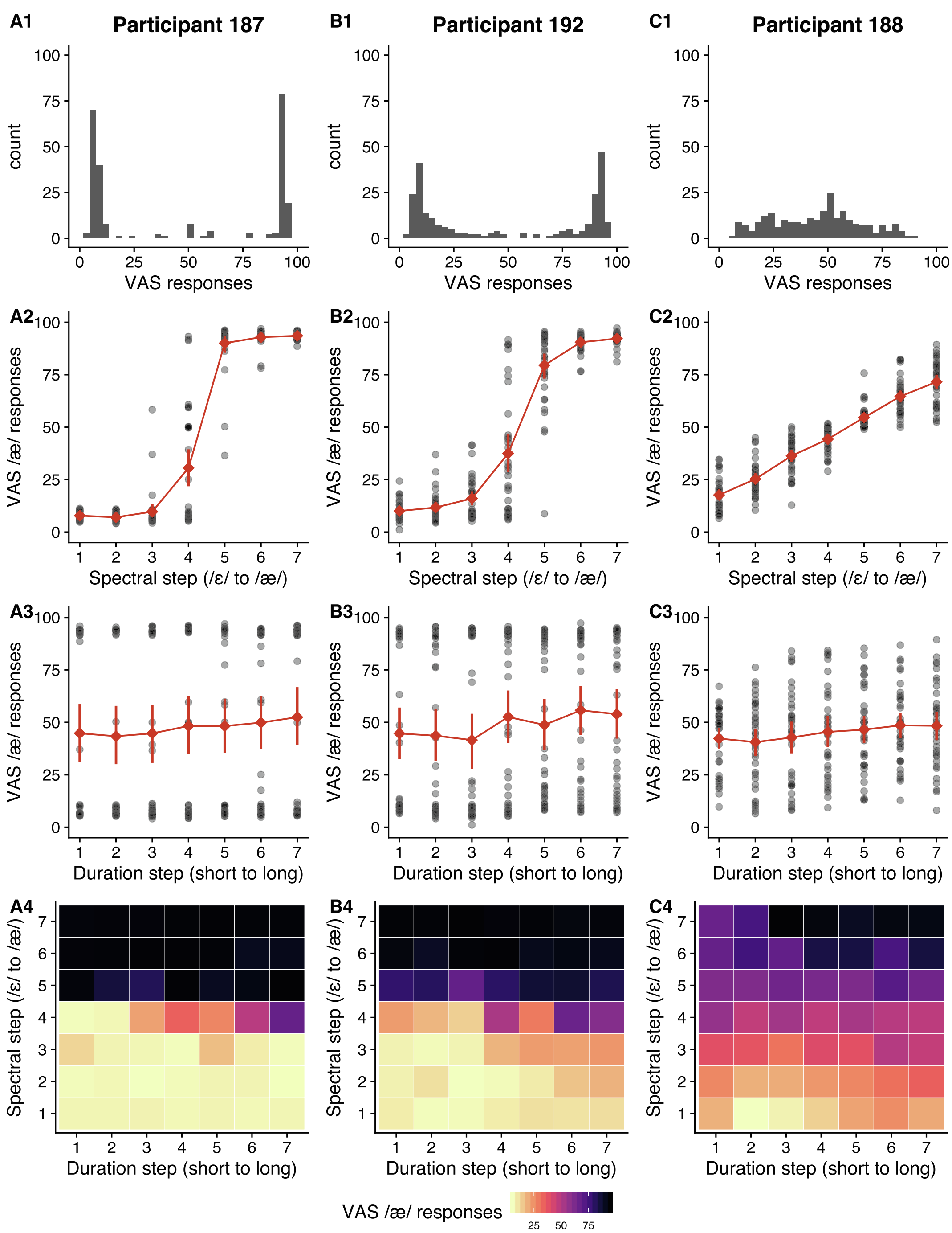

Fig. 5. Visual analog scaling (VAS) responses for three representative listeners (more categorical vs. more gradient). The VAS slope values (gradiency measure) of each representative listener are 132 (Participant 187), 55 (Participant 192), and 10 (Participant 188) in which smaller values indicate more gradient responses.

Table 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	—

(**$p < 0.01$).

siderable variability among listeners, indicating that some listeners made more categorical responses while others made more gradient responses.

Fig. 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 Fig. 5, participants' responses were illustrated by plotting overall VAS responses using histograms (1st row), VAS responses as a function of vowel spectral quality (2nd row) and duration (3rd row). The relative use of each cue was also illustrated using heatmaps (4th row). Specifically, the panels of Participant 187 show that VAS responses were largely categorical clustered around the two endpoints (A1) and the responses were variable at the category boundary as a function of spectral quality (A2) while response patterns were mostly random as a function of duration (A3). Also, the heatmap representation indicates almost exclusive use of spectral quality in vowel categorization (A4). In contrast, the panels of Participant 188 show that VAS responses were more distributed across the entire line (C1) and the responses shifted systematically as a function of spectral quality (C2) and the responses were relatively less clustered around the two endpoints of the VAS scale (C3), which shows a quite different pattern of responses from categorical listeners (e.g., Participant 187). The heatmap also differs from that of Participant 187 in that duration shows some influence on categorization although to much lesser extent than spectral quality (C4).

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. Visual inspection of the heatmaps indicates that these individual differences in gradiency may be associated with the relative use of primary and secondary cues. In the following sections, these differences in categorization gradiency are quantified and compared to other individual difference measures.

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 with one another. Table 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

Table 2
Summary of linear regression model predicting gradiency in phoneme categorization. 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

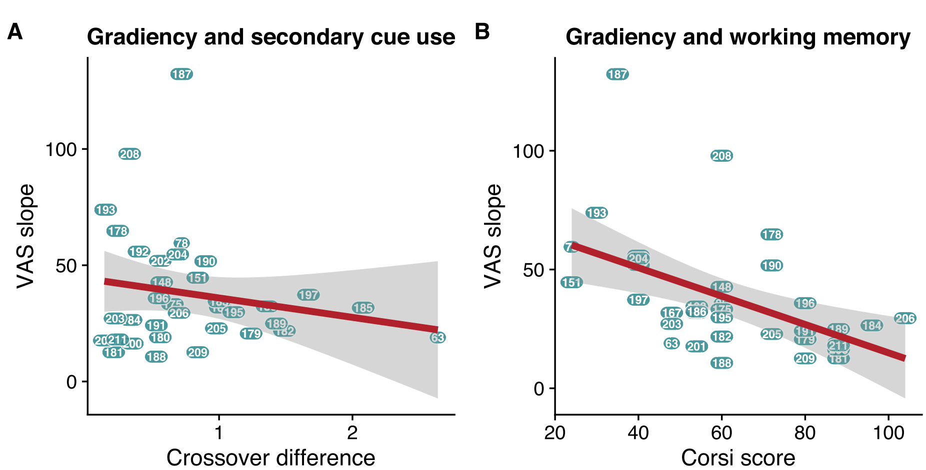

Fig. 6. Categorization gradiency as a function of secondary cue use (A) and working memory (B). The shallower VAS slopes (smaller values) indicate more gradient responses. Higher Corsi scores indicate better working memory capacity.

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.

3.2.3. Relationship between gradiency 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 (Gelman, 2008). Table 2 shows the results of the regression model for categorization gradiency. Each coefficient is the estimated effect when all other predictors are controlled for.

Fig. 6 shows two significant predictors of the regression model. In the model, CoDiff (secondary cue use) significantly predicted the VAS slopes (phoneme categorization gradiency) ($\beta = -16.759$, $t = -2.352$, $p = 0.025$), as shown in Fig. 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 ($\beta = -27.235$, $t = -3.819$, $p < 0.001$), as shown in Fig. 6B. That is, individuals with higher working memory capacity also made more gradient responses in phoneme categorization, in line with Kapnoula et al. (2017).

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 gradiency in phoneme categorization, respectively.

To examine the adaptability of categorization responses, the participants' proportion of /æ/ responses to Test stimuli (short and long vowels with intermediate spectral quality) were analyzed using a mixed-effects logistic regression model. If listeners are adapting, we expect the difference between responses to long and short stimuli (Duration) to increase during Exposure relative to Baseline (Block), namely a Block by Duration interaction. If individual differences measures (VAS, Stroop, Corsi and BCST) are predictive of adaptation, we expect them to modulate this change across blocks, namely three-way interactions between individual differences measures, Block and Duration.

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. Thus, using conservative statistical criteria, the model included all possible random slopes to accurately estimate coefficients and minimize Type I errors despite a loss of statistical power (Barr, Levy, Scheepers, & Tily, 2013; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

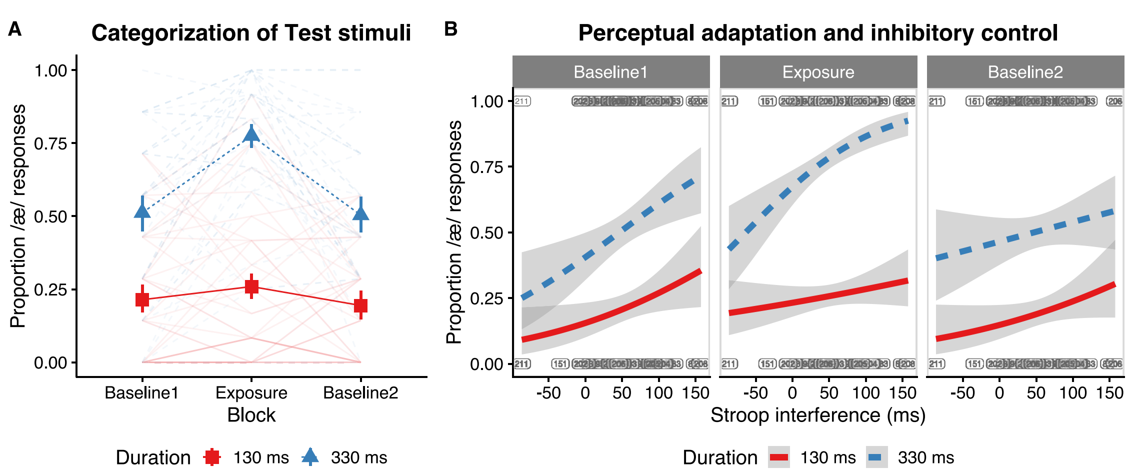

Fig. 7. (A) Proportion of /æ/ vowel responses of the Test stimuli across blocks as a function of short (130 ms) and long (330 ms) vowel durations. Thin lines are individual listeners' Test stimulus responses across blocks. (B) The effect of individual differences in Stroop interference (inhibitory control) on categorization responses for short (130 ms) and long (330 ms) vowel durations across blocks. A higher Stroop interference score indicates low inhibitory control, and each number indicates individual listeners.

Table 3
Summary of fixed-effect coefficients in the logistic regression model of the proportion of /æ/ responses to Test stimuli.[4] Model coefficient estimates (β), standard errors (*SE*), corresponding *z*-values, and *p*-values. Reference levels are indicated in italics.

Predictor	Estimate (β)	*SE*	*z*	*p*
Intercept	−0.637	0.146	−4.342	<0.001
Block_1 (Baseline1 vs. *Exposure*)	−1.372	0.219	−6.243	<0.001
Block_2 (Baseline1 vs. *Baseline2*)	0.659	0.245	2.682	0.007
Duration (130 ms vs. *330 ms*)	−2.365	0.275	−8.595	<0.001
VAS	−0.848	0.386	−2.198	0.027
Stroop	0.927	0.304	3.042	0.002
BCST	0.221	0.305	0.723	0.469
Corsi	0.258	0.342	0.756	0.449
CPT	−0.354	0.289	−1.227	0.219
Block_1 × Duration	1.498	0.435	3.441	<0.001
Block_2 × Duration	−0.812	0.416	−1.949	0.051
Block_1 × VAS	−0.212	0.599	−0.355	0.722
Block_2 × VAS	−0.695	0.659	−1.054	0.291
Block_1 × Stroop	0.126	0.469	0.269	0.788
Block_2 × Stroop	0.631	0.533	1.183	0.236
Block_1 × BCST	−0.611	0.446	−1.370	0.170
Block_2 × BCST	0.487	0.493	0.988	0.323
Block_1 × Corsi	−0.552	0.492	−1.123	0.261
Block_2 × Corsi	−0.381	0.560	−0.680	0.496
Block_1 × CPT	−0.760	0.395	−1.921	0.054
Block_2 × CPT	0.291	0.451	0.645	0.519
Duration × VAS	−1.562	0.724	−2.156	0.031
Duration × Stroop	0.048	0.567	0.085	0.931
Duration × BCST	−0.220	0.565	−0.390	0.696
Duration × Corsi	0.107	0.633	0.170	0.865
Duration × CPT	−1.197	0.532	−2.247	0.024
Block_1 × Duration × VAS	−0.720	1.198	−0.601	0.547
Block_2 × Duration × VAS	−0.187	1.147	−0.163	0.870
Block_1 × Duration × Stroop	1.796	0.924	1.944	0.051
Block_2 × Duration × Stroop	−1.054	0.919	−1.147	0.251
Block_1 × Duration × BCST	0.508	0.877	0.597	0.562
Block_2 × Duration × BCST	−0.970	0.800	−1.213	0.225
Block_1 × Duration × Corsi	−0.871	0.988	−0.882	0.377
Block_2 × Duration × Corsi	0.331	0.922	0.359	0.719
Block_1 × Duration × CPT	−0.135	0.784	−0.172	0.863
Block_2 × Duration × CPT	−0.391	0.717	−0.546	0.585

[4] Since VAS was correlated with Corsi as in Table 1, we also examined adaptation models after removing either VAS or Corsi. However, they did not differ from the results reported here.

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 only on the interactions that are relevant to our research questions outlined above. Specifically, the two-way interaction involving Block and Duration to investigate perceptual adaptation (i.e., more use of duration at Exposure) and the three-way interactions involving Block and Duration and the individual difference measures (VAS and cognitive measures) to investigate whether they predicted perceptual adaptation (i.e., more use of duration at Exposure varying depending on individual differences in gradiency and cognitive abilities).

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 (β = 1.498, z = 3.441, p < 0.001). This is illustrated in Fig. 7A which shows proportion of /æ/ vowel responses to Test stimuli across blocks as a function of short and long vowel durations. 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. Fig. 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.

Individual listeners' Test stimulus responses across blocks (as indicated in thin lines in Fig. 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 these differences. The model found a marginally significant three-way interaction of Block_1 × Duration × Stroop (β = 1.796, z = 1.944, p = 0.051),[3] indicating that greater perceptual adaptation at Exposure may be associated with less inhibitory control. Fig. 7B illustrates this interaction showing a larger increase. Since this finding is rather unexpected and exploratory in nature, it will be explained more in the discussion section in terms of the possibility of less 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

[3] The random intercept-only model showed a significant three-way interaction of Block_1 × Duration × Stroop (β = 1.527, *SE* = 0.691, z = 2.208, p = 0.027), and the random slope model with by-participant random slopes for Block and Duration but no interaction between them also showed a significant three-way interaction of Block_1 × Duration × Stroop (β = 1.690, *SE* = 0.796, z = 2.123, p = 0.033). However, we took a conservative approach by including all possible random slopes (e.g., Barr et al., 2013), which was the model with by-participant random slopes for Block, Duration, and their interaction at the risk of lower statistical power.

(sustained attention) is worth mentioning. That is, the model found a marginally significant interaction of Block$_1$ × CPT ($\beta = -0.760$, $z = -1.921$, $p = 0.054$) indicating that individuals with higher sustained attention may be more sensitive to varying experimental conditions across blocks and a significant interaction of Duration × CPT ($\beta = -1.197$, $z = -2.247$, $p = 0.024$) linking better sustained attention to secondary cue use. However, because there was no three-way interaction between CPT, Duration and Block$_1$, sustained attention was not linked to the magnitude of duration change (i.e., to the magnitude of perceptual adaptation).

4. Discussion

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

4.1. Perceptual adaptation to unfamiliar speech

The current results confirmed previous findings (Idemaru & Holt, 2011, 2014; Lehet & Holt, 2017; Liu & Holt, 2015) that listeners initially adapt to unfamiliar speech patterns at Exposure and subsequently switched their representations back to their long-term category representations when they heard the canonical English pattern at Baseline 2. Both of these results suggest that listeners dynamically adapt to short-term deviations in the input signal while simultaneously maintaining stable long-term representations (Kleinschmidt & Jaeger, 2015). The current results further suggest that the speech perceptual system adjusts to the acoustic consequences of changes in the *relative informativeness* of acoustic dimensions (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Holt & Lotto, 2006; Theodore & Monto, 2019; Toscano & McMurray, 2010). 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 in two ways. The first is that listeners adapted to atypical phonetic categories by up-weighting perceptual reliance on a secondary dimension rather than by down-weighting it as in previous studies of dimension-based statistical learning. The current task also differs from previous work on adaptive changes in cue weights in that listeners had to adapt without any implicit or explicit labeling of the phonetic categories from either the primary cue or explicit feedback (Francis & Nusbaum, 2002; Francis, Baldwin, & Nusbaum, 2000; Harmon, Idemaru, & Kapatsinki, 2019). Recent work by Harmon, Idemaru, and Kapatsinski (2019) on English stop voicing also showed down-weighting of a primary cue (VOT) and up-weighting of a secondary cue (f0) when the primary cue was not informative but the secondary cue was. In their case, the primary cue was made uninformative through explicit feedback on every trial, rather than through being held at relatively ambiguous values. Harmon et al. concluded that reinforcement learning based on trial by trial feedback was the best model of the learning process. However, the fact that we obtained similar results without feedback indicates that feedback may not be necessary to up-weight a secondary cue in all circumstances. In fact, certain kinds of learning from distributional information seem to proceed without any external feedback or labeling at all. One of earliest demonstrations of shifts in speech perception relied on feedback from the lexicon to label ambiguous tokens (Norris et al., 2003). However, subsequent studies have shown that listeners will shift their category boundary based on shifts in the distributions they are exposed to without any lexical labels or explicit feedback (Colby, Clayards, & Baum, 2018; Kleinschmidt & Jaeger, 2015; Munson, 2011; Schreiber, Onishi, & Clayards, 2013). Furthermore, Kleinschmidt, Raizada and Jaeger (2015) compared this unsupervised adaptation to adaptation with implicit feedback on every trial (in the form of available response options) and found that the feedback had no effect on the amount or time-course of learning. Chládková, Podlipský, and Chionidou (2017) directly compared lexically-guided learning (as in Norris et al., 2003) to the same pattern of ambiguous and clear phones in non-words and also found learning in both cases (contrary to the original study), though there were some indications that the learning effect may have been enhanced with the addition of the lexical cues. Colby et al. (2018) also found evidence that lexical cues may enhance learning over purely distributional information, but only for older adults. Thus, the distributions themselves must provide a kind of semi-supervised learning in so far as they differ from long-term expectations (Kleinschmidt, Raizada, & Jaeger, 2015).

Perceptual up-weighting of a secondary cue in response to unfamiliar speech 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 speech production as well as in speech perception under adverse conditions.

Shifting of cue weights to reflect informative dimensions may be a mechanism that is used under other circumstances as well. Increased use of a more informative dimension has been observed when listening in noise (Winn et al., 2013). 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.

4.2. Gradiency and its links to secondary cue use and adaptation

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, Eisner, Schoffelen, Acheson, Hagoort, & McQueen, 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.

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 patterns of phoneme categorization gradiency and perceptual 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. It is also possible that our sample size may not have been big enough to detect a relationship between gradiency and learning.

4.3. The role of cognitive abilities in adaptation and gradiency

This study hinted the possibility that certain patterns of that perceptual adaptation to unfamiliar speech may in part be predicted by individual differences in inhibitory control. In the present study, inhibitory control was the only cognitive measure that was linked to the patterns of adaptation to unfamiliar phonetic categories, although weakly so. This finding is in accordance with previous observations of the potential 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 less inhibitory control showed greater 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. However, recent studies have suggested that reduced inhibitory control can enhance learning performance under some circumstances (Amer, Anderson, & Hasher, 2018; Amer, Campbell, & Hasher, 2016; Weeks, Biss, Murphy, & 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 the ability to process more information, which may be beneficial in certain contexts. For example, in their face recognition study Weeks et al. (2016) found that when participants were shown faces with a name that they were instructed to ignore, individuals with reduced inhibitory control (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, this conclusion should also be taken with caution, as we did not find evidence that those with stronger inhibitory control were less sensitive to duration more generally (lack of interaction between Duration and Stroop). It should further be noted that those with stronger inhibitory control also found the spectrally ambiguous stimuli to be less ambiguous than individuals with weaker inhibitory control (Stroop main effect). Thus there may have been less room for the duration cue to play a role for these listeners. Notably, however, the present finding regarding inhibitory control is preliminary and should not be taken as conclusive. Rather, it may provide a starting point for more systematic investigations of how inhibitory control plays a role in adaptation processes in speech.

In the present study, we also tested individual listeners' ability to maintain alertness over time to control for the effect of sustained attention ability on adaptation performance and speech perception itself. Because the ability to maintain attention should be beneficial to learning (Wickens & McCarley, 2008), we might expect that better ability to sustain attention would be associated with more adaptability. However, we did not find evidence that individual differences in sustained attention were predictive of perceptual adaptation in our task. The study also did not find evidence that categorization gradiency in speech perception is related to listeners' ability to maintain their attention during the task. There was evidence that the ability to sustain attention was related to secondary cue use, however. It is unclear why sustained attention would only be related to secondary cue use but not to gradiency while secondary cue use would be related to gradiency. The relationship between secondary cue use and sustained attention was significant only in the regression model (i.e., the interaction between Duration and CPT in Table 3) but not in the simple correlational analysis (i.e., no significant correlation between CoDiff and CPT in Table 1), so it is possible that sustained attention explains some of the individual differences in secondary cue use that are not already explained by gradiency.

The current finding showed that higher working memory capacity is linked to more gradient processing of speech sounds in the VAS task, as in Kapnoula et al. (2017). Higher working memory capacity may benefit gradient speech perception by facilitating processing and retention of fine-grained within-category differences. On the other hand, perceptual adaptation was not associated with 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. No significant correlations between cognitive ability measures in the study may also indicate that these cognitive factors are not likely related to one another at least in their effects on listeners' perceptual responses.

5. Conclusion

The findings of the present work add to a growing body of research suggesting that listeners are sensitive to short-term changes in distributional information in the speech input. Here we showed that listeners are sensitive to the informativeness of particular cues and flexibly up-weight a secondary cue when a primary cue is temporarily uninformative, even in the absence of any external information such as explicit feedback. This suggests that the relationship between long-term expectations and recent experience may be driving learning.

Furthermore, our results add to the growing interest in 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 of speech sound categories (Kapnoula et al., 2017; Kong & Edwards, 2016), that more gradiency is related to greater secondary cue use and better working memory, and extends previous work to English vowels. The present results also confirm previous work showing that there are considerable individual differences in adaptation patterns (Colby et al., 2018; Schertz et al., 2016) and suggest that these differences may in part be accounted for by individual differences in inhibitory control. 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.

CRediT authorship contribution statement

Donghyun Kim: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing, Investigation, Formal analysis, Visualization. **Meghan Clayards:** Methodology, Writing - review & editing, Supervision, Funding acquisition. **Eun Jong Kong:** Methodology, Writing - review & editing.

Acknowledgements

This work was supported by Social Sciences and Humanities Research Council of Canada [grant number 435-2016-0747] to Meghan Clayards. We thank Jessamyn Schertz, an anonymous reviewer, and the editor (Natasha Warner) for their helpful comments and suggestions. We also thank Morgan Sonderegger for helpful comments on an earlier version of this manuscript.

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. https://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. https://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. https://doi.org/10.1016/j.tics.2016.10.002.

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. https://doi.org/10.1121/1.4922226.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience, 4*(10), 829–839. https://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 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. https://doi.org/10.1121/1.4916265.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language, 94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051.

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. https://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. https://doi.org/10.1121/1.4966677.

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

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. https://doi.org/10.1121/1.2642103.

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

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

Chládková, K., Podlipský, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. *Journal of Experimental Psychology: Human Perception and Performance, 43*(2), 414–427. https://doi.org/10.1037/xhp0000333.

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

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *Journal of the Acoustical Society of America, 144*(3). EL172–EL177 10.1121/1.5052025.

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. https://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.

Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research, 61*(8), 1855–1874. https://doi.org/10.1044/2018_JSLHR-S-17-0392.

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. https://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. https://doi.org/10.1111/lang.12161.

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

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. https://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. https://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. https://doi.org/10.1006/jpho.1997.0052.

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). https://doi.org/10.1371/journal.pone.0063885 e63885.

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. https://doi.org/10.1037/0096-1523.28.2.349.

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

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. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://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*. https://doi.org/10.3389/fnagi.2014.00347.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine, 27*(15), 2865–2873. https://doi.org/10.1002/sim.3107.

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

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition, 189*, 76–88. https://doi.org/10.1016/j.cognition.2019.03.011.

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. https://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. https://doi.org/10.1121/1.2188377.

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. https://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. https://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. https://doi.org/10.1121/1.4765076.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. https://doi.org/10.1038/sj.ki.5002787.

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

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. https://doi.org/10.1080/17470218.2014.964736.

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. https://doi.org/10.1037/xhp0000410.

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In *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. https://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. https://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. https://doi.org/10.1016/j.wocn.2016.08.006.

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. https://doi.org/10.1207/S15324826AN0902_5.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203. https://doi.org/10.1037/a0038695.

Kleinschmidt, D. F., Raizada, R. D., & Jaeger, T. F. (2015). Supervised and unsupervised learning in phonetic adaptation. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the Cognitive Science Society* (pp. 1129–1134). Austin, TX: Cognitive Science Society.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive Psychology, 51*(2), 141–178. https://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. https://doi.org/10.1111/cogs.12413.

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. https://doi.org/10.1016/j.wocn.2014.09.001.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1783–1798. https://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. https://doi.org/10.1037/0033-2909.109.2.163.

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001.

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. https://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. https://doi.org/10.1207/s15516709cog0000_79.

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

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. https://doi.org/10.1177/0963721411429458.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods, 222*, 250–259. https://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. https://doi.org/10.1080/02699206.2016.1233292.

Munson, C. M. (2011). *Perceptual learning in speech reveals pathways of processing* Unpublished PhD dissertation. University of Iowa.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America, 109*(3), 1181–1196. https://doi.org/10.1121/1.1348009.

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

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.

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

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. https://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. https://doi.org/10.1016/j.wocn.2013.03.007.

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. https://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. https://doi.org/10.3758/s13414-015-0987-1.

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

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*(1), 71–80. https://doi.org/10.1016/S0167-6393(02)00094-8.

Schreiber, E., Onishi, K., & Clayards, M. (2013). Manipulating phonological boundaries using distributional cues. *Proceedings of meetings on acoustics, 19*.

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. https://doi.org/10.3766/jaaa.24.7.10.

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review, 1–8*. https://doi.org/10.3758/s13423-018-1551-5.

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. https://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. https://doi.org/10.1177/0956797610384142.

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. https://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. https://doi.org/10.3758/s13423-016-1003-z.

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

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research, 56*(4), 1097–1107. https://doi.org/10.1044/1092-4388(2012/12-0086).

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. https://doi.org/10.1121/1.4883380.