ABSTRACT

Title of Dissertation: INFLUENCE OF SUPPORTIVE CONTEXT AND STIMULUS VARIABILITY ON RAPID ADAPTATION TO NON-NATIVE SPEECH

Rebecca E. Bieber
Doctor of Philosophy 2021

Dissertation directed by: Professor Sandra Gordon-Salant
Hearing and Speech Sciences

Associate Professor Samira Anderson
Hearing and Speech Sciences

Older listeners, particularly those with age-related hearing loss, report a high level of difficulty in perception of non-native speech when queried in clinical settings. In an increasingly global society, addressing these challenges is an important component of providing auditory care and rehabilitation to this population. Prior literature shows that younger listeners can quickly adapt to both unfamiliar and challenging auditory stimuli, improving their perception over a short period of exposure. Prior work has suggested that a protocol including higher variability of the speech materials may be most beneficial for learning; variability within the stimuli may serve to provide listeners with a larger range of acoustic information to map onto higher level lexical representations. However, there is also evidence that increased acoustic variability is not beneficial for all listeners. Listeners also benefit from the presence of semantic context during speech recognition tasks. It is less clear, however, whether older listeners derive more benefit than younger listeners from supportive context; some studies find increased benefit for older listeners, while others find that the context benefit is similar in magnitude across age groups.

This project comprises a series of experiments utilizing behavioral and electrophysiologic measures designed to examine the contributions of acoustic variability and semantic context in relation to speech recognition during the course of rapid adaptation to non-native English speech. Experiment 1 examined the effects of increasing stimulus variability on behavioral measures of rapid adaptation. The results of the study indicated that stimulus variability impacted overall levels of recognition, but did not affect rate of adaptation. This was confirmed in Experiment 2, which also showed that degree of semantic context influenced rate of adaptation, but not overall performance levels. In Experiment 3, younger and older normal-hearing adults showed similar rates of adaptation to a non-native talker regardless of context level, though talker accent and context level interacted to the detriment of older listeners' speech recognition. When cortical responses were examined, younger and older normal-hearing listeners showed similar predictive processing effects for both native and non-native speech.

INFLUENCE OF SUPPORTIVE CONTEXT AND STIMULUS VARIABILITY ON RAPID ADAPTATION TO NON-NATIVE SPEECH

by

Rebecca E. Bieber

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Sandra Gordon-Salant, Chair
Samira Anderson, Co-Chair
Rochelle Newman
Jared Novick
Stefanie Kuchinsky
Catherine Carr, Dean’s Representative

Dedication

For my grandparents, who taught me to embrace change.

Acknowledgements

Thank you to my mentors, Sandy Gordon-Salant and Samira Anderson, who have generously given their time and knowledge to me, and who have been models of success in its many forms. I could not be where I am today without their unwavering support throughout all stages of my graduate studies. Thank you to the members of my committee: Jared Novick, Stefanie Kuchinsky, and Rochelle Newman, whose guidance has shaped both this work and my engagement with the research community.

I am forever grateful to the incredible women in this field who have supported me in various ways throughout my journey, including but not limited to: Carmen Brewer, Kelly King, Anna Tinnemore, Mary Barrett, Julie Cohen, Jaci Schurman, Maddy Buntrock, Talah Wafah, Maureen Shader, Brittany Jaekel, and Calli Yancey. Without you, I could not shine (Sow & Friedman, 2020).

Finally, I would like to thank my family. You have celebrated my highs, and lifted me from my lows. You mean the world to me. Thank you.

This work was supported, in part, by the National Institute on Deafness and Other Communication Disorders (NIDCD) from an institutional training grant (T32DC000046: Co-PIs: C.E. Carr and S. Gordon-Salant) and by the HESP MCM Fund for Student Research Excellence.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACC	anterior cingulate cortex
ANOVA	analysis of variance
ANOM	anomalous predictability
ARHL	age-related hearing loss
BKB	Bamford-Kowal-Bench
CI	cochlear implant
DV	dependent variable
EEG	electroencephalography
ELU	Ease of Language Understanding
ENG	English
ERP	event-related potential
FRA	French
GAMM	generalized additive mixed model
GCM	growth curve model
GLMER	generalized linear mixed effects regression
GN	generalization – new accent
HIN	Hindi
HINT	Hearing-in-Noise Test
HP	high predictability
IEEE	Institute of electrical and electronics engineers
IV	independent variable
JPN	Japanese
KOR	Korean
L1	native language
LMER	linear mixed-effects regression
LP	low predictability
LSWMT	List Sorting Working Memory Task
ML1	multiple native languages
MMN	mismatch negativity
MND	Mandarin

MOCA	Montreal Cognitive Assessment
NE	native English
NIH	National Institutes of Health
NNE	non-native English
NS	native Spanish
OHI	older hearing-impaired
ONH	older normal-hearing
POR	Portuguese
RHHI	Revised Hearing Handicap Inventory
RHT	Reverse Hierarchy Theory
RMS	root-mean-square
RT	reaction time
RUS	Russian
SE	standard error
SNR	signal-to-noise ratio
SPA	Spanish
STD	standard
TG	topic-grouped
TUR	Turkish
UMD	University of Maryland
YNH	young normal-hearing

Chapter 1: Introduction

Introduction

Speech recognition is an active process beginning with the peripheral detection of an acoustic signal and culminating in the comprehension of linguistic information. Sound, in the form of acoustic energy, travels through the external and middle ear systems and is converted into a mechano-electric signal in the cochlea. From there, the signal travels along the eighth cranial nerve to the auditory cortex via the auditory brainstem, with final processing of acoustic signals taking place in the primary auditory cortex. While the mechanisms of the acoustic/mechano-electric transduction of the auditory signal are fairly well understood, the conversion of sound to meaning is yet to be as clearly defined.

Generally, successful speech perception is thought to be facilitated by both ‘bottom-up’ and ‘top-down’ processes. The lower-level (i.e. ‘bottom’) information comprises fine-grained acoustic detail, whereas higher level processes (i.e. ‘top’) include a listener’s rich knowledge of the language. Though the exact processes underlying speech recognition remain debated, speech perception is clearly a remarkably robust, necessarily flexible process. This flexibility is critical for successful speech recognition under varying conditions. Speech recognition performance can be modified by factors inherent to the listener, talker, or listening environment.

Listener-related factors

The relative success or failure of a spoken communication encounter can be highly influenced by factors related to the listener. Most relevant to the present project, advanced age is often associated with poorer speech recognition outcomes (Dubno et al., 1984), especially in combination with age-related hearing loss (ARHL). The effects of aging and ARHL on speech recognition as they relate to this project will be discussed in detail below.

Certain individual characteristics may mediate some of the detriments of aging and ARHL. When the target speech signal is altered in some way (see further discussion of signal variability below), it appears that a listener's prior experience with that form of signal alteration can greatly benefit that listener. For example, Gordon-Salant and Friedman (2011) showed that, in a group of older blind adults, regular recreational listening to pre-recorded materials (i.e. audiobooks) at accelerated play-back rates correlated with higher recognition scores for time-compressed sentences. This benefit of prior experience has also been shown for recognition of unfamiliar speech patterns (Mcgarr, 1983; Tjaden & Liss, 1995) and non-native accents (Hanulíková & Weber, 2012; Porretta et al., 2016; Scott & Cutler, 1984; Sumner & Samuel, 2009; Weber et al., 2014; Witteman et al., 2013).

Another individual factor that may influence speech recognition is the individual's cognitive capacity. The relationship between cognitive abilities and speech recognition ability has been explored extensively, though there are still significant gaps in knowledge. The field of "cognitive hearing science"

(Arlinger et al., 2009) has grown rapidly in recent years, with numerous studies investigating the relationship between various cognitive and linguistic domains and speech perception, particularly in challenging environments. The Ease of Language Understanding (ELU) model (Rönnberg et al., 2008), a model of speech recognition, posits that cognitive functions play an important role in facilitating speech understanding, with an emphasis on the importance of executive functions, especially working memory.

On the whole, cognitive skills falling under the domain of executive function tend to be associated with speech recognition outcomes. These may include, among others, working memory, inhibition, and attentional control (Jurado & Rosselli, 2007). Working memory represents the capacity to store and manipulate information, and consistently emerges as a significant predictor of individual performance for speech recognition, including speech in noise (Akeroyd, 2008; Anderson, White-Schwoch, et al., 2013; Füllgrabe et al., 2015) as well as other forms of challenging speech. Although working memory is not consistently a significant predictor of speech recognition performance for younger listeners with normal hearing thresholds, it does contribute to performance for listeners in middle age and older (Füllgrabe & Rosen, 2016). The utility of the processes associated with working memory (i.e. processing and storage) in the context of speech recognition is clear: a listener must retain and manipulate acoustic-phonetic and linguistic information in order to successfully participate in spoken conversation, which necessarily involves a rapid rate of incoming sensory information.

The ELU model has been updated in recent years to include consideration of other aspects of executive function that are critical for speech recognition, including inhibition and attentional control (Rönnberg et al., 2013). Inhibition and attentional control are contrasted in that attentional control reflects the ability to appropriately allocate processing resources to stimuli of interest, while inhibition is the process by which the undesired allocation of processing resources to non-relevant cues is prevented. The hypothesized role of inhibitory mechanisms in speech recognition is in restricting information that is irrelevant to the target from taking up resources that would be used for processing the target speech. Measures of inhibition correlate with recognition of speech in the presence of competing talkers and in challenging environments (Dey & Sommers, 2015; Janse, 2012a; Sommers & Danielson, 1999).

A connection between attentional mechanisms and successful recognition of challenging speech is also logical. Listeners must be able to successfully attend to relevant acoustic/phonetic features of the auditory input in order to successfully recognize target speech. There is evidence that selective attention modulates the neural representations of target versus masking speech in a speech-in-noise task (Golumbic et al., 2013). Interestingly, however, attentional mechanisms may interact with the specifics of the recognition task. For example, auditory perceptual learning (see below) is reduced when participants are explicitly directed to attend to the specific acoustic features of the stimuli (McAuliffe & Babel, 2016), especially in participants with better attention-switching control (Scharenborg et al., 2015). Scharenborg et al. (2015) suggest that those with greater attention-switching control can make more use of bottom-up acoustic cues and rely less on higher-level lexico-semantic information, and are thus hindered in accomplishing lexically guided learning.

Stimulus-related factors

Speech stimuli can be modified by factors relating to the listening environment (presence of background noise or reverberation) or to the talker (accelerated speech rate or non-native speech). The experiments in this project focus on one type of signal alteration: presence of a non-native accent.

Speech produced by non-native talkers is often associated with poorer or more effortful speech recognition performance as compared to native speech (Goslin et al., 2012; Porretta & Kyröläinen, 2019; Wade et al., 2007). It appears that this effect may be greater for older adults (Gordon-Salant et al., 2010a, b), though some researchers do not find a clear difference between age groups (Ferguson et al., 2010). A non-native accent results from a combination of the segmental (i.e. phoneme-level), subsegmental (i.e. acoustic-level), and suprasegmental (i.e. word or phrase-level) features of a non-native talker’s first language and the second language in which they are speaking (Flege, 1988). Segmental substitutions might include exchanging one sound for another, or different realizations of the same target sound (i.e. changes in nasalization, or vowel height). Suprasegmental changes can include differences in fundamental frequency range, syllable stress, and overall prosody, and subsegmental features include alterations in vowel durations. These changes lead to alterations of the speech signal, and can cause listeners difficulty when mapping lexical meaning onto an acoustic signal that may not align with the expected acoustic features of an unaccented production. In addition to containing altered acoustic features, non-native productions can be more acoustically variable than speech produced by native talkers (Wade et al., 2007; Xie & Jaeger, 2020).

Early models of speech recognition assumed that listeners performed some sort of normalization to variable and/or challenging stimuli such that variability would not impede recognition, but the current understanding is that listeners do not recognize speech in this exemplar fashion, but rather make use of the information inherent to the variability to aid recognition (Pisoni, 1997). While a non-native accent alters the acoustic characteristics of the speech signal, the presence of a non-native accent is also a type of indexical feature that may serve as a cue to the speaker's identity. Indexical features include information about the talker, including their gender, age, language background, affective state, etc. Work by Pisoni and colleagues (Nygaard et al., 1994; Nygaard & Pisoni, 1998; Palmeri et al., 1993; Pisoni, 1997) has documented that listeners retain not only the lexical aspects of target speech, but also the indexical properties of the speakers they have heard, and can use this information to improve speech recognition. These findings have informed the understanding of the speech recognition process, with acoustic, lexical, and indexical properties of speech all contributing to the mapping of incoming signals to flexible mental representations. Similarly, Cai et al. (2017) proposed that listeners use a "speaker-model" of speech recognition, in which information about the talker (i.e., indexical information) is used to help predict speech and interpret meaning to facilitate recognition.

Even in the absence of signal distortions or audibility limitations, natural speech is still highly variable: the acoustic properties of the same lexical item will not be identical from iteration to iteration, both across and within individual speakers. Yet, listeners (especially young adult listeners with normal hearing) appear to be unaffected by this variance, with recognition of spoken communication continuing fairly smoothly. Further, when the speech signal is altered, most young adult listeners are able to overcome an initial decrease in speech recognition ability after a short period of exposure, improving their recognition for this challenging speech over time.

Adaptation and perceptual learning

The improvement in speech recognition over time is known as adaptation, and is understood to be a manifestation of perceptual learning in the auditory domain. As defined by Goldstone (1998), perceptual learning occurs when there are 'long-lasting changes to [the] perceptual system that improve [the] ability to respond to [the] environment and are caused by this environment'. Reverse Hierarchy Theory (RHT) is a model of perceptual learning proposed by Ahissar and Hochstein (2004), originally developed in relation to the visual system, and later applied to the auditory system (Ahissar et al., 2009). RHT suggests that perception strongly relies on higher processing levels and moves in a top-down fashion, where the lower levels of perception are only accessed (via ‘backward search’) as needed. For example, RHT would posit that a listener perceives a word as a whole unit, rather than as the sum of its acoustic features. When challenging speech is encountered, listeners may not be able to perform this high-level whole-unit perception, and must rely on the low-level acoustic features for perception of words and phrases. Over the course of perceptual learning, listeners are thought to adjust their internal high-level representations of lexical items, decreasing reliance on the low-level detail. This allows for improvements in perception and recognition over time. The high-level adjustment of internal representations has been demonstrated behaviorally (Dahan et al., 2008), suggesting that listeners perform flexible on-line adjustments when communicating, using their knowledge of the target speaker’s characteristics (i.e. language background).

There is extensive evidence of perceptual learning in the auditory domain [see Kraljic and Samuel (2009) for a review]. Auditory perceptual learning occurs for simple auditory stimuli, such as tones differing in spectral (Amitay et al., 2005; Irvine et al., 2000; Menning et al., 2000; Wright & Sabin, 2007) or temporal (Wright et al., 1997; Wright & Sabin, 2007) characteristics. Auditory perceptual learning has also been documented for unfamiliar phonetic contrasts, both those that exist in languages other than the listener’s native language (L1; Lively et al., 1993) and those that have been artificially constructed for research purposes (Norris et al., 2003; Reinke et al., 2003; Tremblay et al., 1997, 2001). Similarly, auditory perceptual learning has been documented for complex, sentence-length acoustic stimuli, for both naturalistic (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Nygaard & Pisoni, 1998) and artificial signal alterations (Adank & Janse, 2010; Janse & Adank, 2012; Maye et al., 2008). Auditory perceptual learning is evident in behavioral as well as objective electrophysiologic measures (Atienza et al., 2002; Menning et al., 2000; Reinke et al., 2003; Romero-Rivas et al., 2015; Song et al., 2008; Tremblay et al., 1997, 2001).

Rapid adaptation to a non-native talker can occur even after a very brief period of exposure (Clarke & Garrett, 2004; Sidaras et al., 2009), and is thought to comprise the early, fast stage of perceptual earning. Perceptual learning is understood to include fast and slow processes, which result in both short- and long-term improvements in speech recognition. This distinction is detailed by Atienza and colleagues (2002), who measured event-related potential (ERP) indicators of perceptual learning at four 12-hour time intervals following a frequency discrimination training session. They described the fast and slow processes as “acquisition and consolidation of perceptual learning, respectively”, and documented both immediate and delayed changes to ERP responses, supporting a temporal separation between the two processes. In this paper, the focus is on the immediate, fast process, which will be referred to as “rapid adaptation” rather than the long-lasting changes to the perceptual system, which are often targeted by auditory training programs.

Adaptation, learning, and generalization

For the purposes of this project, ‘auditory training’ refers to protocols involving pre-testing and post-testing, separated by an intervention that is intended to modify performance in the post-test relative to the pre-test. Auditory training often occurs over several sessions. ‘Rapid adaptation’ refers to the early period of on-line perceptual adjustment, comprising change in performance over a period of time as short as a few minutes. In many cases, auditory training can elicit an initial period of rapid adaptation. In addition to comparing pre- and post-tests, some researchers report performance *during* training, which in this paper may be referred to as the adaptation period. For some of the literature reviewed in this introduction, pre-test or training data are considered to reflect rapid adaptation (even if the authors do not label them as adaptation data) if they contain the listener’s first exposure to that type of stimulus in the experiment and enough detail to allow for examination of performance over a fine time scale.

Auditory training and learning tasks often include a test of generalization, evaluating performance on different stimuli and tasks from those included in the training (Wright & Zhang, 2009). These generalization tests can clarify which components of the learned task can generalize to stimuli with different acoustic features or to different lexical items. For example, a series of papers that examined perceptual learning for ambiguous phonemes had contrasting findings: Kraljic and Samuel (2006) found that listeners generalized their learning to both new talkers and new phonemes, while Eisner and McQueen (2005) found that learning did not extend to new speakers. Maye et al. (2008) examined learning for an artificial accent involving vowel-lowering and found that learning was specific to the type of distortion introduced: learning did not generalize to words in which vowels had been artificially raised.

Generalization is particularly of interest for the evaluation of rehabilitative auditory training protocols, which ideally benefit listeners not only during training, but also in improvements to daily communication. Generalization can be evaluated in a number of configurations, ranging from generalization across talkers (i.e. identical stimuli produced by different talkers) and across stimulus types (i.e. different stimulus types produced by familiar talkers), to generalization across tasks (i.e. different outcome measures). Researchers have also investigated whether or not training on cognitive tasks generalizes to improvements in speech in noise performance (Ingvalson et al., 2017; Wayne et al., 2016), or vice versa (Ferguson et al., 2014; Ferguson & Henshaw, 2015).

In this project, the following generalization types are considered: generalization across different talkers for familiar and unfamiliar sentences and generalization across non-native accents. Generalization across accents necessarily includes generalization to a new talker. Prior studies of adaptation to non-native speech have documented both generalization to new talkers with the same accent (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009) and to new talkers with different, unfamiliar non-native accents

(Baese-Berk et al., 2013; Bieber & Gordon-Salant, 2017). Reports vary as to the time course of generalization. Studies such as those of Baese-Berk et al. (2013) and others (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Scharenborg et al., 2015; Scharenborg & Janse, 2013) suggest that generalization occurs immediately after a period of training. Others indicate that generalization occurs on a slower time course than perceptual learning (Zaltz et al., 2020) or is dependent on the duration of training (Wright et al., 2010).

Perceptual Adaptation and Aging

Aging can have a detrimental effect on many aspects of the speech recognition process. Aging is associated with ARHL; the loss of sensory acuity can contribute to decreased speech recognition ability. Even in the absence of hearing loss, aging is associated with deficits in auditory processing. Age-related reductions in auditory temporal processing ability can exacerbate challenges with recognition of temporally altered speech, including non-native speech (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2010a, 2010b). Aging can also have an effect on cognitive capacity and mechanisms that contribute to successful speech recognition (Füllgrabe et al., 2015). It is well documented that aging has a detrimental effect on cognitive function across domains including processing speed and working memory (Lipnicki et al., 2017; Salthouse, 1990, 2004). One domain that does not decline with increasing age is vocabulary; indeed, vocabulary size appears to continue to grow into older adulthood (Salthouse, 2004). This may serve as a protective mechanism for some older adults, as vocabulary has been shown to be predictive of learning-consistent behavior (Colby et al., 2018).

Cognitive predictors of degraded speech recognition and/or adaptation appear to differ between age groups (Dey & Sommers, 2015; Ingvalson et al., 2017), though contrary results have been reported (Colby et al., 2018). Ingvalson and colleagues (2017) also found that hearing acuity interacted with some cognitive factors in predicting non-native speech recognition. The behavioral findings of age-related differences in cognitive predictors are supported by literature indicating that older adults recruit different cortical regions for speech recognition than do younger listeners (Eckert et al., 2008; Erb & Obleser, 2013). For example, Erb and Obleser (2013) found that, in a study of perceptual adaptation to vocoded speech, older adults showed a persistent elevation and decreased dynamic range of activation of the anterior cingulate cortex (ACC) for both distorted and clear conditions, a prefrontal region associated with cognitive control (Stevens et al., 2011), while younger adults showed a clearer distinction in ACC activation between clear and distorted speech conditions.

Older adults maintain an ability to adapt to distorted speech signals over a short period of exposure. However, the process of adaptation may differ between younger and older adults. There are numerous reports of different patterns of adaptation between younger and older listeners in studies examining adaptation to non-native speech (Adank & Janse, 2010; Bieber & Gordon-Salant, 2017), time-compressed speech (Manheim et al., 2018;

Peelle & Wingfield, 2005), ambiguous phonemes (Scharenborg & Janse, 2013), and speech in noise (Karawani et al., 2016). These pattern-wise differences do seem to vary across studies. Some authors report ‘unlearning’ (i.e. a decline in performance following an initial increase) in their younger but not older listeners (Karawani et al., 2016; Scharenborg & Janse, 2013), while others report a plateau and slight unlearning in older listeners (Adank & Janse, 2010). Colby et al. (2018) report some unlearning over the course of perceptual learning, which differs by stimulus type rather than listener age. Regarding rates of learning, some find a steeper or more rapid rate of adaptation in younger versus older listeners (Adank & Janse, 2010; Bieber & Gordon-Salant, 2017; Peelle & Wingfield, 2005). In contrast, Manheim et al. (2018) report a steeper rate of adaptation for older hearing-impaired listeners compared to either younger or older normal-hearing listeners, though these authors note that this may have been driven by differences in starting level (i.e. the older hearing-impaired listeners had more room to improve). This report of greater learning for participants at lower starting levels is not uncommon (Banks et al., 2015; Ferguson et al., 2014). However, some authors report no differences in the rates and patterns of adaptation between younger and older listeners (Erb & Obleser, 2013; Gordon-Salant et al., 2010; Neger et al., 2014). In a study of adaptation to time-compressed speech, Peelle and Wingfield (2005) reported no age differences when the two groups were matched on starting performance level, but did find an age effect on the pattern of adaptation when the two groups were listening to the same degree of time compression: younger listeners showed a greater degree of early learning than older listeners.

The inconsistencies in the findings regarding age effects on perceptual learning for degraded or challenging speech may arise from a number of sources. For example, consideration of hearing loss varies widely across studies. In some cases, listeners with hearing impairment are treated as a separate participant group (Bieber & Gordon-Salant, 2017; Gordon-Salant, Yeni-Komshian, Fitzgibbons, et al., 2010; Manheim et al., 2018), while other “older” participant groups include listeners with and without hearing impairment (Adank and Janse, 2010, Erb and Obleser, 2013, Janse and Adank, 2012, Neger et al., 2014, Scharenborg and Janse, 2013). For a true understanding of the mechanisms underlying rapid adaptation in older listeners, it is critical to consider hearing sensitivity as a factor, as hearing impairment may have a differential or exacerbating effect on recognition of distorted or challenging speech (Sommers, 1997). For example, Manheim and colleagues (2018) report both age-related and hearing loss-related effects on the magnitude of rapid adaptation to time-compressed speech, but only hearing loss effects on rate of adaptation.

Other aspects of study methodology may affect the outcomes regarding the influence of age on rapid adaptation, such as the ways in which performance and adaptation are measured. Studies have used outcome measures such as percent correct repetition or transcription, reaction times for lexical decisions or other forms of comprehension, signal-to-noise ratios

(SNRs) for targeted performance levels, eye-tracking, and event-related potentials. Adaptation can be quantified by examining either the rate or magnitude of change in performance, and the ways in which these can be quantified may differ. Age effects may be apparent in some methodologies but not others, depending on precisely what is being probed, and what component of the adaptation process is examined. For example, Adank and Janse (2010) found that younger and older adults had a similar magnitude of adaptation, but showed differences in the time course of learning. Task parameters, such as whether or not the task is speeded, may also influence findings (Janse & Jesse, 2014), with age effects more likely to emerge under time constraints. There is a clear need for research that combines and compares methodologies in order to facilitate a deeper understanding of the adaptation process.

Factors Facilitating Perceptual Adaptation: Semantic Context

Semantic context and speech recognition

Availability of semantic context supports speech recognition: words in isolation or in sentences with a low degree of semantic context are recognized less accurately than words presented in semantically meaningful sentences (Kalikow et al., 1977; Miller et al., 1951; Nittrouer & Boothroyd, 1990). As a sentence unfolds, a listener makes use of the contextual information within the sentence to generate predictions about upcoming words. A sentence that is rich in semantic information is described as highly constraining, i.e. the number of potential words that may occur within that context is constrained. In contrast, a weakly constraining sentence may contain any of a large number of words. This scale of weak/strong sentential constraint is related to a distinction of target word predictability. While ‘constraint’ refers to the sentence frame (i.e. how much does it reduce the number of options for upcoming words?), ‘predictability’ refers to the characteristics of the specific word (i.e. how likely is it that the word will occur in this sentence?). In the behavioral literature, a corpus of high-predictability (HP) and low-predictability (LP) sentences (Bilger et al., 1984; Kalikow et al., 1977) is often used, with differences in performance between HP and LP performance operationalized as a ‘context benefit.’ This relative improvement in behavioral recognition can also be tested with electrophysiological measures.

The event-related potential (ERP) component most commonly used to examine effects of semantic context on speech recognition is known as the N400. The N400 component is a negative-going potential that typically occurs between 300-500 ms, and reflects ease of lexical access (Lau et al., 2008). Lexical access involves the mapping of a target auditory signal to an internal lexical representation. In other words, lexical access is the complex process by which listeners assign meaning to sound. The N400 component is theoretically present in response to any stimulus, but the magnitude of the negative deflection depends on the ease with which that item was accessed. Lexical access can be facilitated or hindered in various ways, but the presence/absence of supportive, congruent, or semantic context is the canonical example of an experimental manipulation employed to generate an N400 *effect* (i.e. a relative difference in N400 *component* amplitude between conditions). The N400 effect is often visualized as a difference wave between two conditions varying in predictability or congruence.

Updated models of speech recognition acknowledge that sentential context and lexical information (Ganong, 1980) play a role in resolving lexical ambiguity (Gaskell & Marslen-Wilson, 2001). Interestingly, it appears that the benefit of lexico-semantic information is modulated by the quality of the target speech, and the presence/absence of signal distortions (Aydelott et al., 2006; Aydelott & Bates, 2004; Goy et al., 2013; Straus et al., 2013). For example, Goy et al. (2013) examined the semantic context benefit using three different forms of signal distortion, including low-pass filtering, time-compression, and concurrent 12-talker babble. The benefit of semantic information in facilitating recognition was compared between an undistorted condition and all three types of signal distortion; the facilitation score was reduced in all conditions, with no differences across distortion type. However, the benefit of context was not completely eliminated in the acoustically distorted conditions.

Semantic context and learning

Availability of lexical-semantic information has also been shown to influence perceptual learning and rapid adaptation to unfamiliar or challenging speech (Davis et al., 2005; Maye et al., 2008; Norris et al., 2003). For example, Davis and colleagues (2005) performed a series of experiments examining perceptual learning for noise-vocoded speech, a manipulation that degrades much of the spectral information in speech but preserves temporal characteristics. Listeners were trained on various types of vocoded sentences and were subsequently tested with standard English vocoded sentences. The training conditions included vocoded versions of standard English sentences, semantically anomalous (but syntactically intact with real words) sentences, and Jabberwocky (syntactically intact with non-real content words) sentences. The goal was to see what degree of lexical information was most effective for facilitating generalization to vocoded standard English sentences. For example, the Jabberwocky sentences contain an intermediate level of lexical information, as the sentences contain both real words and non-words, with syntactic structure preserved. They found that listeners who trained on sentences containing lexical information (standard English, semantically anomalous, Jabberwocky) all showed greater learning than those who trained with non-words or who did not train at all. The presence of lexical information in the training stimuli appears to have been critical for learning of spectrally distorted stimuli.

Semantic context and aging

While the benefit of lexico-semantic information for speech recognition is well documented, the influence of listener age on this context benefit is not as clear. In studies of the contextual benefit which use auditory stimuli, various findings suggest that older listeners can benefit less (Federmeier et al., 2002, 2003; Schurman et al., 2014), equally (Dubno et al., 2000; Sheldon et al., 2008; Wingfield et al., 1994), or more (Goy et al., 2013; Pichora-Fuller et al., 1995; Sommers & Danielson, 1999) than younger listeners from the presence of semantic context during otherwise challenging listening situations. Some also find a greater context benefit for older listeners with hearing impairment as compared to their normal-hearing peers (Janse & Jesse, 2014; Pichora-Fuller et al., 1995). There is behavioral evidence that older listeners may even demonstrate an over-reliance on semantic context (Rogers & Wingfield, 2015), and an inability to inhibit lexical-semantic knowledge during perception (Mattys & Scharenborg, 2014) or recall (Hartman & Hasher, 1991), factors that may lead to misperceptions. In these cases, the presence of a constraining context could be detrimental to speech recognition for older listeners. It is challenging to discern the factors underlying these different findings across studies.

First, the various operational definitions of a ‘context benefit’ should be considered. The classic behavioral studies (Dubno et al., 2000; Pichora-Fuller et al., 1995; Sheldon et al., 2008) all calculate the context benefit by examining the relative tolerance for signal distortion (i.e. vocoding bands, noise level) in the different context conditions. For example, Pichora-Fuller et al. (1995) compared performance with HP and LP sentences over a range of signal-to-noise ratios (SNRs), and concluded that older listeners with and without hearing impairment benefitted from context at a great number of SNRs than the YNH listeners.

Other researchers measure the context benefit by including conditions with several levels of context (often representing various levels of cloze predictability) and testing whether context level interacts with age in predicting the outcome variables such as keyword repetition (Benichov et al., 2012), or latency and amplitude of event-related potentials (ERPs; Federmeier et al., 2003, 2010; Federmeier & Kutas, 2005). Aydelott et al. (2006) and Goy et al. (2013) both report some age effects on the context benefit, but these seem to vary depending on presence and type of background noise. Janse and Jesse (2014) report that the level of the available context predicts outcomes as measured by reaction time, but not accuracy, in a group of older listeners. One consideration in study methodology is ensuring a similar baseline performance level, so that the magnitude of the context benefit can be measured under equivalent conditions. This potential issue is discussed by Dubno and colleagues (2000), who conclude that older and younger adults derive equivalent benefit from the presence of context under listening conditions allowing for similar baseline performance. Collectively, it is clear that there are differences in findings depending on study methodology.

The electrophysiological examinations of semantic processing in older adults typically document a negative effect of aging: the N400 effect, an indicator of predictive processing, is delayed and reduced in older listeners (Federmeier et al., 2002, 2003; Federmeier & Kutas, 2005; Kutas & Iragui, 1998; Wlotko et al., 2012, though note that hearing thresholds are not collected in these studies). The age effect seems to originate from older adults' limited ability to benefit from a strongly constraining semantic context in order to facilitate predictive processing. In other words, these studies find age effects for the conditions containing rich semantic information, but not for those with low or anomalous predictability. For example, Federmeier and Kutas (2005) found that younger and older adults had a similar magnitude ERP response to target words presented in a weakly constraining sentence context, but that older adults showed a smaller context-facilitated reduction in N400 component magnitude to words in strongly constraining sentence contexts than did younger listeners. There is also electrophysiologic evidence that older listeners recruit different neural regions when processing semantic information, despite similar behavioral performance (Lacombe et al., 2015). Age differences in the processing of semantic information may be evident in online (i.e. EEG) but not offline (i.e. sentence repetition) measures.

The addition of a memory load may also contribute to the emergence of age differences in the benefit of context (Pichora-Fuller et al., 1995; Schurman et al., 2014; Wingfield, 1996). For example, Wingfield (1996) evaluated the benefit of increasing levels of context for both older and younger listeners in two paradigms: one including a working memory load and one without. The study showed no interaction of aging and context for the condition with no working memory component, but found that older adults gained less benefit from context when working memory was implicated. A working memory load can be introduced either by including an explicit working memory task during the experiment, or by varying the level of degradation to the target signals under the assumption that signal degradation increases the reliance on working memory for speech recognition, as posited by the ELU model (Rönnberg et al., 2013). These differences could contribute to the contrasting findings across different studies; a comparison of findings between methodologies with consistent stimulus/task parameters is warranted.

Factors Hindering Perceptual Adaptation: Stimulus Variability

Stimulus variability and speech recognition

Stimulus variability is also known to have an effect on speech recognition, with greater variability typically associated with poorer performance. For example, speech recognition is higher for a list of stimuli produced by a single talker than by multiple talkers (Bent & Holt, 2013; Mullennix et al., 1989; Sommers et al., 1994). Another factor that may contribute to acoustic stimulus variability is a mismatch between the talker's native language and the language in which they are speaking. Across talkers, non-native speakers are more variable in their speech productions than native speakers (Wade et al., 2007). A recent study by Xie and Jaeger (2020) examined within-talker acoustic variability in American English productions of native Mandarin speakers. They quantified acoustic variability by examining category means and separation as well as magnitude and orientation of cue dispersion for vowels, closures, and bursts. The analyses showed that within-talker variability was increased for some but not all components of Mandarin-accented English. This increased acoustic variability may contribute to the difficulties experienced when listening to non-native speech.

Stimulus variability and learning

Variability has also been explored as a mechanism by which to facilitate improved recognition of non-native speech, and has gained interest as a way to promote generalization of learning for non-native speech (i.e. improved performance with unfamiliar talkers). In the literature, manipulations of variability have been operationalized as single vs. multiple talkers (Bradlow & Bent, 2008; Lively et al., 1993), blocking vs. randomization of stimulus features during training (Tzeng et al., 2016), adjustments of vowel formant distributions (Wade et al., 2007), strength of non-native accent (Witteman et al., 2014), single vs. multiple talker language backgrounds (Alexander & Nygaard, 2019; Baese-Berk et al., 2013), and consistency of signal distortion (Golomb et al., 2007; Gordon-Salant & Fitzgibbons, 2004).

Early studies by Pisoni and colleagues (Bradlow et al., 1999; Lively et al., 1993; Logan et al., 1991) documented the benefit of variability in training listeners to perceive non-native contrasts. Listeners who were trained on stimuli recorded by multiple talkers showed greater generalization to new talkers compared to those who trained on one talker. These findings have been extended to a recent line of work that provides evidence that exposure to multiple non-native talkers (with either similar or dissimilar native languages) improves recognition for unfamiliar talkers and/or unfamiliar accents (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Janse & Adank, 2012; Sidaras et al., 2009). Some suggest that this benefit is related to an increased flexibility of internal phonetic-lexical representations, as a result of exposure to the systematic variations in non-native speech. As argued by Baese-Berk et al. (2013), certain speech sounds and patterns may be more susceptible to deviations in production by non-native talkers, and listeners may be able to utilize the high level of variability to develop greater flexibility in recognition. Others suggest that the benefit of multiple talkers is related to an intrinsically higher likelihood that talkers encountered during exposure/training are more likely to be acoustically similar to the stimuli with which generalization is tested (Xie & Myers, 2017). Other studies have shown that generalization of rapid learning is contingent on the perceptual features of the generalization stimulus (Borrie et al., 2017; Reinisch & Holt, 2013).

One recent study (Alexander & Nygaard, 2019) examined the relative benefits of training with single accents versus multiple accents in facilitating recognition of non-native speech. Overall, they found that there was no significant difference in performance with an unfamiliar non-native talker between groups who had been trained with a single accent or multiple accents, if the test talker’s accent differed from those present in the training stimuli. There was weak evidence for a benefit of training with multiple talkers, but only for one of the generalization conditions.

While training with multiple talkers appears to be beneficial in facilitating generalization to new talkers, it is also worth considering how the level of stimulus variability influences the immediate, rapid adaptation process, which is a closer analogue to everyday interactions with unfamiliar talkers. Generally speaking, higher levels of stimulus variability appear to contribute to shallower and/or less linear patterns of adaptation (Bradlow & Bent, 2008; Tzeng et al., 2016; Wade et al., 2007; Witteman et al., 2014). The overall magnitude of adaptation is not diminished by higher levels of variability (Tzeng et al., 2016; Witteman et al., 2014), although there are reports to the contrary (Bradlow & Bent, 2008; Wade et al., 2007). These disparate findings may relate to the nature of the experimental variability and the types of low-variability comparisons. Tzeng et al. (2016) and Witteman et al. (2014) both compared conditions containing similar acoustic information. Tzeng and colleagues (2016) examined the effects of presentation order during training, while holding the overall amount of acoustic variability in the entire experiment stable across groups. A cohort of young, normal-hearing listeners heard English sentences produced by 4 different native Spanish speakers (2 male, 2 female). The stimuli were blocked by talker, by sentence, or were randomly presented, but listeners in all conditions heard the same stimuli. Witteman et al. (2014) compared two conditions utilizing stimuli by the same talker, which varied in that talker’s consistency of production. In both studies, listeners showed a similar or greater magnitude of adaptation in the conditions with greater variability.

In contrast, studies comparing conditions with distinctly different acoustic characteristics show a limited magnitude of adaptation for the higher variability conditions (Bradlow & Bent, 2008; Wade et al., 2007). For example, Wade et al. (2007) examined the effects of non-native variability in vowel production on adaptation and perceptual learning in young adults. The stimuli were synthetically constructed vowels based on the vowel formant distribution patterns of either native English speakers or native Spanish speakers producing target English speech. The distributions were based on the means and standard deviations of these two talker populations. Listeners who heard sounds falling within the variability distribution of non-native talkers (L1: Spanish) did not show adaptation to the stimuli, while listeners who heard tokens with native-like vowel formant distributions did improve their recognition.

Together, these findings suggest that variability in the form of acoustic similarity/difference can have a detrimental effect on both the rate and overall magnitude of rapid adaptation. The assumption is that since higher levels of variability lead to increased difficulty in speech recognition, the adaptation process would also be hindered by a more acoustically variable stimulus. It remains unclear whether detriments imposed by high levels of acoustic variability can be offset by the introduction of supportive semantic context. Despite these negative effects on the time course of adaptation, however, greater acoustic variability in the adaptation stimulus does appear to facilitate greater generalization to new, unfamiliar stimuli.

Stimulus variability and aging

To date, there have been limited examinations of the effect of stimulus variability as imposed by different degrees of non-native accent on older listeners. However, a small number of studies examined these effects using other forms of signal alteration (Golomb et al., 2007; Sommers, 1997) with older adults. Sommers (1997) explored the effects of stimulus variability on monosyllabic speech recognition with older listeners with and without ARHL. This study varied three forms of acoustic-phonetic information: number of talkers, speech rate, and overall amplitude. When the manipulated dimension was number of talkers, both age and ARHL interacted with variability in predicting word recognition, suggesting additive effects of both aging and ARHL for this variability manipulation. When speech rate was varied, there was a significant interaction with ARHL but not age, and when amplitude was varied, there was a significant interaction with age but not ARHL. Overall, these findings suggest that age and ARHL effects are not consistent over different forms of variability.

Golomb et al. (2007) examined the effects of variability for a time-compressed speech signal, where the variability manipulation involved the consistency and/or regularity with which the speech was time compressed. The researchers found no age differences in amount of learning for time-compressed speech in conditions that were varied based on degree and type of interruptions to the signal. Perhaps age effects in overall recognition depend on the type of stimulus, or the form of variability manipulation.

Sommers' (1997) variability manipulation of number of talkers includes changes to spectral and indexical information, whereas changes in speech rate/time compression are more limited in terms of acoustic variability.

Summary and Overall Goal

The long-term goal of this research is to increase the understanding of intrinsic and extrinsic factors that contribute to a listener's ability to adapt to unfamiliar or challenging speech. Understanding the mechanisms underlying this rapid adaptation to challenging speech and facilitating adaptation in older adults is critical in informing comprehensive hearing health care, including audiologic rehabilitation programs and counseling. This project consists of a series of experiments designed to examine two extrinsic factors that are predicted to influence rapid adaptation: acoustic variability in the speech stimulus, and presence of supportive semantic context. Stimulus variability has been shown to be detrimental to speech recognition, but exposure to variability may contribute to improved recognition of unfamiliar stimuli. Presence of semantic context has been shown to facilitate speech recognition under degraded listening conditions, but some questions remain about older listeners' ability to benefit from contextual information, and whether it is contingent on hearing status. A thorough examination of these factors and their specific contributions to rapid adaptation, rather than overall recognition, has not been completed.

The central hypothesis is that rapid adaptation is facilitated by a listener's flexibility in mapping acoustic input to lexical knowledge, but that aging involves a reduction in this flexibility. For older listeners, therefore, higher levels of variability are predicted to diminish adaptation, while higher levels of stimulus context should support increased adaptation. The two factors are expected to interact such that the benefit of context is greatest for intermediate levels of acoustic variability. This project combines both behavioral and electrophysiologic measures of speech processing, allowing for a parallel examination of the processes contributing to speech recognition during the course of rapid adaptation to non-native speech stimuli.

Chapter 2: Study 1

Assessing the influence of stimulus variability on rapid adaptation and generalization to unfamiliar stimuli

Introduction

Recognition of degraded or challenging speech

When communicating under sub-optimal conditions, most young adult listeners with normal hearing (YNH) are able to quickly adapt to the speech signal, improving their recognition with additional exposure. This process, termed rapid adaptation, has been well-documented in YNH listeners for speech that has been time compressed (Dupoux & Green, 1997), noise-vocoded (Davis et al., 2005; Hervais-Adelman et al., 2008; Hervais-Adelman et al., 2011; Huyck et al., 2017), or produced by non-native talkers (Alexander & Nygaard, 2019; Banks et al., 2015; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009). One stimulus-related factor that appears to play a role in rapid adaptation and learning for degraded speech signals is the degree of variability present in the stimulus. Stimulus variability has been defined in different ways throughout the literature. Controlled variability manipulations can include differences in spectral, temporal, or indexical features of the stimuli, or a combination of the above. Classic literature regarding stimulus variability has demonstrated that speech recognition is poorer for lists of stimuli containing tokens from multiple talkers as compared to a single talker (Mullennix et al., 1989; Sommers, 1997). The manipulation in number of talkers potentially produces variability in both acoustic (i.e. a broader distribution of acoustic features) and indexical (i.e. multiple talker identities, multiple talker genders) domains.

Stimulus variability and learning/adaptation

While stimulus variability is known to have a detrimental effect on overall speech recognition, a growing body of work suggests that variability during auditory training is beneficial in facilitating generalization of learning. This was highlighted in early studies regarding perceptual learning of non-native phoneme contrasts. Pisoni and colleagues (Bradlow et al., 1999; Lively et al., 1993, 1994; Logan et al., 1991) published a series of studies in which native speakers of Japanese were trained to distinguish the liquid phonemes /l/ and /r/, a contrast that is not present in Japanese, but is present in American English. In a 1993 study, Lively and colleagues found that listeners who were trained on lists containing productions of /l/ and /r/ by multiple talkers showed improvements during training, but also generalized their learning to both unfamiliar words and unfamiliar talkers. A group of listeners who were only trained on a single talker showed improvement during training, but did not generalize to either an unfamiliar talker or unfamiliar words produced by the familiar talker. Interestingly, the multiple-talker group showed consistent improvements throughout the three-week course of training, while the single-talker group only improved during the first half of training.

High-variability training has also been explored as a mechanism for facilitating improvements in the recognition of non-native speech, on the word or sentence level. For example, Bradlow and Bent (2008) found that native speakers of English showed greater generalization to speech produced by non-native talkers (i.e. foreign-accented speech) after exposure to multiple talkers from a shared language background compared to just one talker. The hypothesized mechanism underlying this benefit relates to systematic variability in the realization of non-native speech. Talkers with a shared native language (L1) will have similar features when speaking in a second language (e.g., leading to the phenomenon of Spanish-accented English), and there are certain features of speech that are more susceptible to alterations when produced by non-native talkers. These facts led researchers to hypothesize that exposure to multiple non-native talkers would benefit listeners in facilitating a higher likelihood of recognition of non-native speech due to experience with tokens encompassing a range of potential alterations to the signal.

This benefit of systematic variability was tested by Baese-Berk et al. (2013), who exposed native English listeners to multiple non-native talkers with various L1s with the goal of facilitating improved recognition of an unfamiliar talker from an unfamiliar language background. They found that, compared to a group of native English listeners who had been exposed to non-native talkers from a single language background, listeners who had been exposed to the multiple-language talkers showed higher performance with an unfamiliar talker. The passive training paradigm employed by Baese-Berk et al. (2013) was designed to expose listeners to a high degree of acoustic variability in the training stimulus. Unlike the protocols employed by

Pisoni and colleagues (Lively et al., 1993; Logan et al., 1991) and Bent and Bradlow (2008), there was no change in the number of talkers. Rather, the presumption was that a group of non-native talkers with unique L1s would have a higher degree of variability in the acoustic features of the target speech as compared to a group of native talkers, or non-native talkers with a shared L1 (see Wade et al., 2007 and Xie & Fowler, 2020 for analyses of the acoustic variability present in non-native speech).

The protocol used by Baese-Berk et al. (2013) includes the presentation of stimuli that are blocked by talker: an identical list of 16 sentences was presented 5 times, each by a different talker. All listeners heard the stimuli in this blocked manner. Other researchers have explored variability manipulations in which the overall acoustic and indexical information remains constant over the course of the experiment, but the orders of presentation differed. Tzeng and colleagues (2016) presented their participants with listening conditions including stimuli that were blocked by sentence, speaker, or were totally randomized. The goal was to evaluate the conditions under which listeners would show evidence of perceptual learning for Spanish-accented English, examining both the time course of learning and the magnitude of generalization. The results of the study suggested that the time course of rapid adaptation varied depending on the variability conditions. There was strong early improvement across conditions, but only the Random group continued to improve reliably over the entire course of the training. Further, when evaluating generalization to unfamiliar talkers with Spanish accents, transfer of learning was only seen for the group who heard randomly presented stimuli.

Variability, learning, and aging

The protocols described above all involved testing with young, college-aged adults with normal hearing. There have been no prior studies examining the effect of stimulus variability on rapid adaptation or perceptual learning for non-native speech in older adults. However, some tentative hypotheses can be drawn based on related literature. One study specifically examining the effects of stimulus variability and listener age on *recognition* (but not learning) indicated that older adults were more affected by some forms of variability than others, and that the presence of ARHL exacerbated some of the detriments imposed by aging (Sommers, 1997). However, all stimuli in this study were produced by native English talkers.

When rapid adaptation to a non-native accent is tested in older and younger adults, there is typically no age-related difference in the magnitude of adaptation (Adank & Janse, 2010; Bieber & Gordon-Salant, 2017; Gordon-Salant, et al., 2010), though younger adults appear to show a more linear pattern of adaptation (Adank & Janse, 2010; Bieber & Gordon-Salant, 2017). This literature is quite limited at present, and no studies have included an explicit variability manipulation, such as number of talkers, in addition to the presence of non-native speech.

Collectively, the limited literature suggests that when performing rapid adaptation to an unfamiliar accent and/or talker, older adults may be more susceptible to increases in variability as compared to YNH listeners, with variability-related detriments being exaggerated in these listeners. An examination of the effects of increasing talker-related variability on rapid adaptation to non-native speech in older adults is warranted.

Summary and hypotheses

The goal of this study is to assess the extent to which rapid adaptation and generalization to non-native speech stimuli are influenced by the degree of variability present in the stimulus, and whether these effects are modified by aging and hearing loss. The working hypothesis is that higher degrees of stimulus variability will not affect adaptation in younger listeners, but will result in a slower rate and reduced magnitude of adaptation for older listeners. Because older listeners can be differentially affected by a non-native accent in overall recognition of non-native speech (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2010b), it is expected that learning for this form of signal degradation will occur on a slower time scale under the assumption of finite processing resources (Kahneman, 1973) and slowed cognitive/perceptual processing for older adults (Salthouse, 2004). Alternatively, there is evidence for a greater magnitude and/or faster rate of learning for listeners who have a lower baseline performance (Banks et al., 2015; Henshaw & Ferguson, 2014; Manheim et al., 2018); in this study, starting performance will be equated as best as possible using background noise.

Talker-independent generalization is expected to be greater for conditions with higher variability, for both younger and older listeners.

Younger adults are expected to generalize to talkers with unfamiliar non-native accents in high-variability conditions, consistent with the findings of Baese-Berk et al. (2013). The expectations regarding accent-independent generalization in older adults are less clear. The theorized benefit of exposure to systematic variability in non-native speech would suggest that all listeners would generalize to an unfamiliar accent. However, if the older adults demonstrate a reduced rate of learning, accent-independent generalization may be reduced or absent for these listeners, regardless of hearing status. Transfer is thought to occur on a later time scale than perceptual learning, at least for non-speech stimuli (Ortiz & Wright, 2010), and different forms of generalization are also thought to occur at different points in time. Thus, if older adults are slowed in their perceptual learning, generalization may be slowed as well, and may not be captured in the present protocol. Considering these factors and the known age-related slowing in cognitive processing (Salthouse, 2004), different patterns of generalization are expected between younger and older listeners. The limited literature regarding ARHL and rapid learning of non-native speech suggests that ONH and OHI listeners will demonstrate similar patterns of generalization (Bieber & Gordon-Salant, 2017).

Method

Participants

The participants for this study included three groups of 20 listeners: younger listeners with normal hearing (YNH), older listeners with normal hearing (ONH), and older listeners with hearing impairment (OHI). Group sizes were determined via *a priori* power analysis targeting a power level of 0.8 at an alpha level of 0.05 (Westfall et al., 2014). Normal hearing was defined as pure-tone thresholds of ≤25 dB HL from 250-4000 Hz; hearing impairment was defined as at least a mild hearing loss (i.e. 30 dB HL) at all octave frequencies from 2-8 kHz. Listeners reporting a history of middle ear disease or neurologic impairment were excluded from participation. Prior to testing, all listeners also completed a screening test for mild cognitive impairment (MoCA, Nasreddine et al., 2005). Listeners who did not fit the hearing-related criteria or have a score ≥ 26 on the MoCA were excluded from participation. Additionally, all listeners were required to have at least a high school education, to speak only American English as their first language, and to report no languages other than English spoken in the home before the age of 7. Details about the listeners can be seen in Table 2.1

Table 2.1. Characteristics of participants in Experiment 1. HFPTA = high frequency pure-tone average (0.5, 1, 2, 4 kHz).

Listener group	**n**	**Age in years** M(*sd*)	**HFPTA in dB HL** M(*sd*)	**MOCA score M (*sd*)**	**Stroop effect in ms** M(*sd*)
YNH	20	20.52 (*1.46*)	6.17 (*4.59*)	28.50 (*1.24*)	157.45 (*88.16*)
ONH	20	68.71 (*3.84*)	15.08 (*5.34*)	27.95 (*1.43*)	369.44 (*148.58*)
OHI	20	73.49 (*4.96*)	40.25 (*8.31*)	28.05 (*2.78*)	363.22 (*234.71*)

Stimuli and procedure

Stimuli. Stimuli for this experiment included BKB/HINT sentences (Bench et al., 1979) produced by two male native English (NE) and 17 non-native English (NNE) speakers, all targeting a standard American English dialect. The BKB/HINT sentences are simple, declarative sentences that include ~4-6 keywords per sentence. The NNE speakers’ native languages (L1s) include Hindi, French, Japanese, Korean, Mandarin, Portuguese, Russian, Spanish, and Turkish. The majority of the NNE stimuli were obtained from SpeechBox (formerly the Online Speech/Corpora Archive and Analysis Resource (Bradlow, n.d.)). One additional talker was recruited from the UMD community and made recordings in the Hearing Research laboratory. Recordings were made using a Shure MS48 microphone and a Marantz Professional PMD661 Handheld Solid State Recorder. Stimuli were spliced from the raw recordings using Adobe Audition 2018, and equalized for root-mean-square (RMS) amplitude using Praat version 6.0.47 (Boersma & Weenink, 2019). A 1000 Hz calibration tone that was equal in RMS level to the sentence stimuli was generated in Praat.

The NNE talkers had foreign accent strength ratings falling between 4.14-6.76 (mean 5.59) on a scale of 1-9 (Atagi & Bent, 2011, 2013) and had equivalent intelligibility ratings to each other and to the NE talkers, as assessed by pilot testing with 10 YNH listeners. Stimulus lists were constructed to have similar, high intelligibility levels (i.e. ≥90% correct identification in quiet by YNH listeners). Stimuli were presented in multitalker babble (6 talkers, male and female native English speakers) at signal-to-noise ratios (SNRs) of +7 dB SNR for the normal-hearing (YNH, ONH) groups, and +10 for the OHI group, with a signal level of 65 dB SPL. These SNR levels were set after a series of pilot tests targeted at setting an equivalent starting performance level (~40-60% correct) across listener groups in the various conditions. Additional details about the talkers and their characteristics can be found in Table 2.2.

Table 2.2. Characteristics of talkers in Experiment 1. L1 = Native language; ENG = English; SPA = Spanish; JPN = Japanese; HIN = Hindi; FRA = French; KOR = Korean; MND = Mandarin; POR = Portuguese; RUS = Russian; TUR = Turkish; Gen: Generalization

Talker ID	Database	L1	Experimental Condition	Accent rating (x/9)	Avg. Intelligibility (all sentences used)
E1M	Speechbox	ENG	C1	1.11	100%
E5M	Speechbox	ENG	C1	1.42	100%
S1M	Speechbox	SPA	C2	5.89	96%
662	Speechbox	SPA	C2	6.24	97%
J1M	Speechbox	JPN	C3	5.86	95%
J2M	Speechbox	JPN	C3	5.48	94%
UMD1S	UMD	SPA	C4	5.74	99%
838	Speechbox	SPA	C4	6.76	99%
663	Speechbox	SPA	C4	4.48	98%
839	Speechbox	SPA	C4	5.48	97%
841	Speechbox	HIN	C5	4.72	97%
F2M	Speechbox	FRA	C5	6.31	94%
652	Speechbox	SPA	C5	5.4	100%
837	Speechbox	SPA	C5	6.48	98%
K1M	Speechbox	KOR	Gen – Unfamiliar	5.77	95%
M1M	Speechbox	MND	Gen – Unfamiliar	4.14	96%
139	Speechbox	POR	Gen – Unfamiliar	5.72	99%
845	Speechbox	RUS	Gen – Unfamiliar	5.28	97%
661	Speechbox	TUR	Gen - Unfamiliar	5.36	100%

Procedure. The experiment included five conditions, each assessing rapid adaptation (i.e. the expected improved speech recognition within a list of 30 sentences) and two forms of generalization (one list of 10 sentences per generalization form). The conditions are as follows: C1: Single native English (ENG) talker; C2: Single non-native English talker (SPA); C3: Single non-native English talker (JPN), C4: Three NNE talkers (shared L1 - SPA); C5: Three NNE talkers (multiple L1s – SPA, HIN, FRA). Order of conditions (C1 through C5) was randomized, with the talkers used for adaptation and generalization alternated and counterbalanced in order to minimize potential talker effects. In each condition, all listeners completed an adaptation phase consisting of 30 trials. This was immediately followed by a test of generalization to a familiar accent, where listeners were tested on an unfamiliar talker who shared their L1 with the talker(s) heard during adaptation. In C5 (Multiple L1s), this unfamiliar talker was a native speaker of Spanish. This was then followed by a test of generalization to an unfamiliar talker with an unfamiliar accent. Participants were given breaks of 5-10 minutes in between each condition; cognitive measures and questionnaires were completed in between the main speech recognition tasks (adaptation and generalization phases). The speech recognition trial structure (i.e. presentation, repetition, feedback, second presentation) was identical in the adaptation and generalization phases. The study protocol is schematized in Figure 2.1.

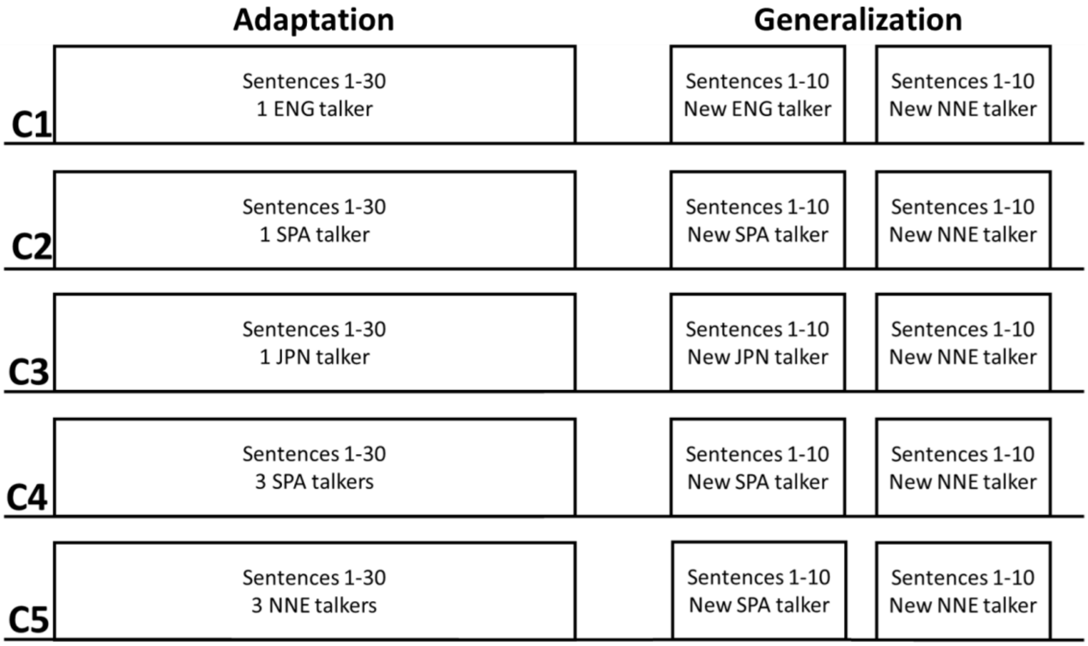


Figure 2.1. Study protocol for this Experiment. In each condition, listeners completed the adaptation phase, which was immediately followed by two forms of generalization testing with unfamiliar talkers: familiar accent (center), and new accent (right). Order of conditions was randomized within and across listener groups. Talkers heard during the Generalization-new accent phase were randomized within and across listener groups. C1: Single ENG; C2: Single SPA; C3: Single JPN; C4: Multiple SPA; C5: Multiple language backgrounds; NNE: Non-native English talkers (including Korean, Mandarin, Portuguese, Russian, and Turkish).

In each trial, participants heard a sentence and repeated it back to the best of their ability. Responses were scored for keywords correct; responses were recorded as .wav files and stored anonymously for confirmation of scoring. If the participant repeated the sentence correctly, they received visual feedback (“Correct!”) on the computer monitor. Following their response, all participants heard the original sentence spoken again by the same talker while the sentence text was presented on a computer screen. This presentation of a clear stimulus following the distorted stimulus has been shown to facilitate lexically guided rapid adaptation (Davis et al., 2005). Stimulus presentation and collection of responses were controlled using E.Prime 2.0 software.

All listeners additionally completed cognitive and linguistic measures, including vocabulary knowledge (NIH Cognitive Toolbox, Weintraub et al., 2013), inhibition (Stroop task, Stroop, 1935), attention (Trail-Making Task, Reitan, 1971), executive function (Card Sort Task, Weintraub et al., 2013), and working memory (List Sorting Task, Weintraub et al., 2013), as well as the Language History Questionnaire 3.0 (Li et al., 2019).

Stimulus variability. The arrangement of talkers and accents in each condition was designed to result in different levels of stimulus variability per condition and to allow for a series of planned comparisons: C1 vs C2 (effect of non-native accent); C2 vs C4 (effect of number of talkers with the same L1 [single vs. multiple]); C2 vs C3 (effect of single talker L1); C4 vs C5 (effect of multiple talker L1 [uniform vs varied]). The conditions examined in this experiment comprise several forms of variability manipulation, including changes to both the acoustic and indexical features of the target stimuli.

The comparison of C1 (single ENG) and C2 (single SPA) is designed to examine the effect of native vs non-native speech. As described by Wade et al. (2007), non-native English talkers display greater acoustic variability in their production of vowel phonemes than do native English speakers. Xie and Jaeger (2020) explored this claim in more detail; measurements similar to those made by both Wade et al (2007) and Xie and Jaeger (2020) were made on the present stimulus sets. However, this single-talker comparison, as well as that between C2 (single SPA) and C3 (single JPN) does not involve any change to the number of talkers or languages represented within each adaptation list.

The C2 (single SPA) vs C4 (multiple SPA) and C4 (multiple SPA) vs C5 (multiple L1s) comparisons include several forms of variability. For C2 vs C4, the indexical features are more variable, as the stimuli are produced by three different talkers in C4 vs one talker in C2. This should result in increased variability of acoustic features, as each talker has their own idiosyncratic features in producing Spanish-accented English. Similarly, the C4 vs C5 comparison is expected to result in increased stimulus variability, as the three talkers in C5 all have different language backgrounds. The acoustic alterations contained within the Spanish-accented English heard in C4 are expected to be relatively similar across talkers, as compared to three different accents heard in C5. This difference should also impose greater indexical variability, as listeners may identify the different accents across talkers.

Acoustic analyses

In order to describe the acoustic features and examine acoustic variability across conditions, acoustic analyses were performed on all sentences that were heard during the adaptation conditions. The following parameters were analyzed from each talker’s recordings to examine components of the acoustic variability: average sentence duration, speech rate (syllables/second), and vowel formant measures. All acoustic analyses were performed using Praat. For analyses of vowel features, phoneme boundaries for the vowels contained in each word sentence were marked using Praat. A summary of the acoustic measures can be found in Table 2.3.

Table 2.3. Acoustic features of stimuli in Experiment 1

Condition	Duration mean (*sd*)	Syllable rate mean (*sd*)	Variability of /i/	Variability of /ɪ/	Separability of /i/ from /ɪ/	Separability of /ɪ/ from /i/
Single ENG	1.59 (*0.18*)	3.77 (*0.6*)	151.09	212.04	490.42	503.90
Single SPA	1.89 (*0.32*)	3.28 (*0.63*)	288.64	191.16	321.40	252.07
Single JPN	1.83 (*0.37*)	3.71 (*0.3*)	152.46	205.10	197.73	233.01
Multiple SPA	2.15 (*0.37*)	3.03 (*0.64*)	288.77	184.33	309.91	212.39
Multiple L1s	2.11 (*0.36*)	3.18 (*0.6*)	142.18	230.09	334.10	255.53

Sentence duration and speech rate. Sentence duration (in seconds) and speech rate (syllables/second) were calculated for each sentence, and are shown in Figure 2.2.

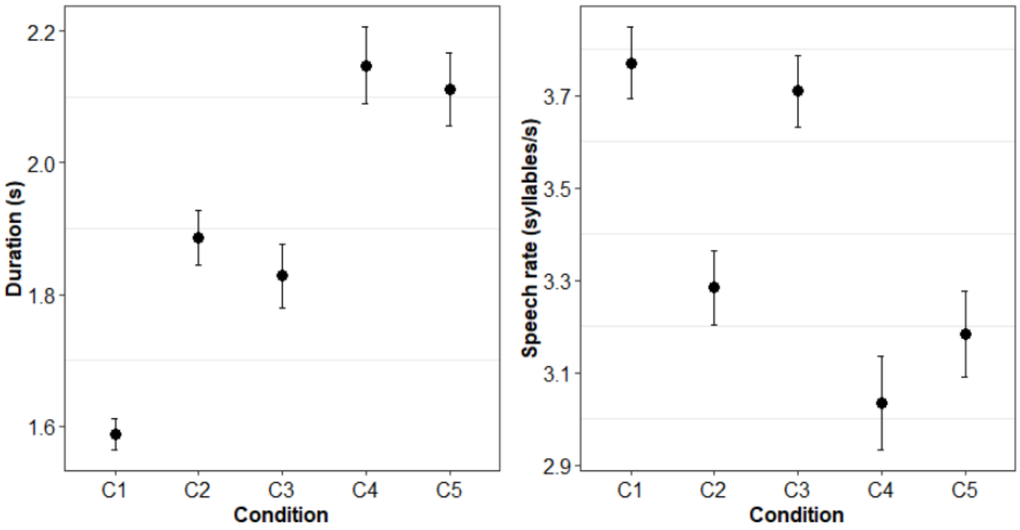


Figure 2.2. Sentence duration and speech rate for stimuli heard during adaptation conditions. Error bars reflect standard error. C1: Single ENG; C2:

Single SPA; C3: Single JPN; C4: Multiple SPA; C5: Multiple language backgrounds.

Sentence durations were entered in a linear mixed effects regression (LMER) model examining the effect of condition on sentence duration, including a random intercept of sentence. This analysis showed that the sentence durations were shorter in C1 than in all other conditions (*p*<.001, all comparisons). Additionally, sentence durations in the multiple NNE talker conditions (C4 and C5) were even longer than sentence durations in the single NNE talker conditions (C2 and C3; *p*<.05, all comparisons). A similar LMER analysis modelling speech rate showed that the sentences in all NNE conditions were characterized by slower speech rates than in C1 (*p*<.001, all comparisons), with the exception of C3. Comparisons of the single vs multiple NNE talker conditions showed that speech rate did not differ significantly between C2 and either C4 or C5 (*p*=0.35, .096, respectively), though C3 did have a significantly faster rate than both C4 (*p*<.001) and C5 (*p*<.01).

Category separability and magnitude of dispersion. Separability is a measure of vowel features that is described by Xie and Jaeger (2020) in a recent study of acoustic variability in non-native-accented English speech. The measure is derived by calculating the mean F1 and F2 values for the phoneme pair of interest, and then comparing that mean to each token of the opposite member of the pair (See Xie & Jaeger 2020 for details on calculation). As a representative index, the category separability was measured for all /i/ and /ɪ/ phonemes present in the recorded sentences used in the experimental adaptation conditions. The dispersion value is utilized by Xie and Jaeger (2020) as an index of variability. Similar to the separability measure, it is generated by comparing each token to the mean value of its own vowel category. Formant measures were extracted for each vowel using Praat; the F1 and F2 values for all /i/ and /ɪ/ stimuli are visualized in Figure 2.3.

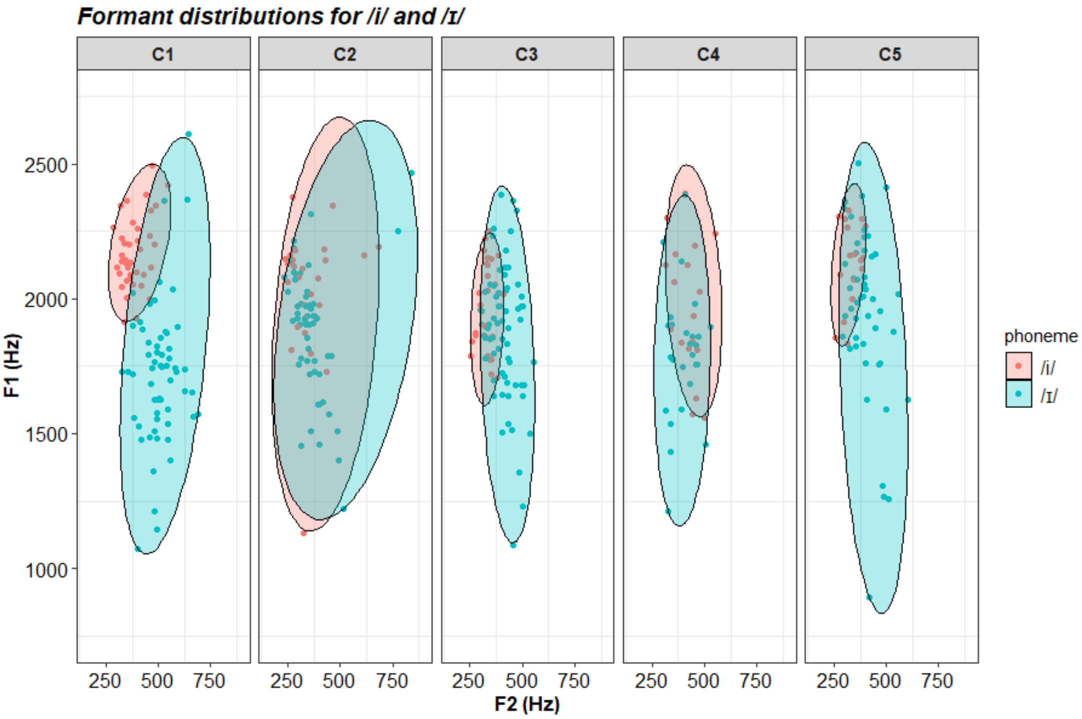


Figure 2.3. Distribution of F1 and F2 values for /i/ (red) and /ɪ/ (blue) phonemes produced within sentences heard during adaptation. Each point represents a single production. C1: Single ENG; C2: Single SPA; C3: Single JPN; C4: Multiple SPA; C5: Multiple language backgrounds.

The category separability and dispersion measures are reported in Table 2.3, along with the measures of duration and syllable rate. Collectively, the measures suggest that the acoustic features differed across conditions, both in terms of absolute values and in terms of variability, though the patterns were not consistent across measures. For example, sentence duration showed greater variability as indexed by standard deviation in the NNE conditions as compared to the single ENG condition, but this was not seen for speech rate. Similarly, /i/ was noted to have greater variability in some but not all of the NNE conditions as compared to the single ENG condition. This aligns with the findings of Xie and Jaeger (2020), who reported greater acoustic variability for non-native speakers for some vowel types but not others.

Statistical analyses: speech recognition and adaptation

The speech recognition analyses center around four main outcome measures: time course of adaptation, magnitude of adaptation, generalization to unfamiliar talkers with familiar accents, and generalization to unfamiliar talkers with unfamiliar accents.

Time course of adaptation

In order to model the non-linear time course patterns of adaptation, Generalized Additive Mixed Models (GAMMs; Hastie & Tibshirani, 1987; Wood, 2006) were utilized. GAMMs make use of non-parametric smoothing functions that include a series of underlying basis functions in order to fit curves to a dataset. GAMM analysis allows for a detailed understanding of which predictor variables cause the trajectories of the dependent variable (DV) to differ over time and at which time points the trajectories differ from one another, which is relevant to the study of rapid adaptation to challenging speech. GAMMs can also be used to account for the autocorrelation inherently present in much time-series data. However, interpretation of GAMMs is not as straightforward as linear mixed-effects regression, and inclusion of higher-order interactions is challenging, particularly with categorical predictors. Of particular note, the strategies for significance testing in GAMMs often differ from those for GLMER and growth curve analysis (GCA), an extension of GLMER. Depending on the research questions and the variable coding strategies, the significance of p-values within the model summary may or may not be meaningful; inspection of fitted data in conjunction with likelihood ratio testing is recommended for interpreting the model findings and determining significance of individual terms (Sóskuthy, 2017, 2021).

For each experimental comparison in this study, a 3 (Group) x 2 (Condition) model was specified using ordinal-coded difference smooths. This coding strategy allows for the construction of models that can distinguish between intercept-level and slope-level differences in two curves. Thus, these models were used to examine whether the manipulation of talker type resulted in overall changes in performance level, changes in the slopes of performance over time, or both.

For the GAMM adaptation analyses, the ONH listener group was used as the reference level, and contrast-coded variables were constructed to evaluate the interactions between Condition and Group for YNH vs ONH (i.e. effect of age) and ONH vs OHI (i.e. effect of hearing loss). All GAMM analyses for time course of adaptation included random smooths for token, and random reference/difference smooths were included for random smooths of subject by condition (Sóskuthy, 2021). Bonferroni correction was used with ordinal coded models, as each comparison is represented twice within the same model in order to examine both intercept-level and slope-level effects (Sóskuthy, 2021). Separate GAMM models including a combined-factor variable, and tensor product interaction smooths were also evaluated to examine the effects of individual characteristics on the time-course of rapid adaptation. Tensor product interaction terms allow for evaluation of the non-linear interaction of the individual measure with the Group and Condition variables in predicting speech recognition scores. For all GAMM models, a weighted binomial distribution was utilized.

Magnitude of adaptation

To examine the magnitude of adaptation, a derived value for relative improvement over the course of adaptation was calculated for each listener by averaging performance over the first and last five trials, subtracting performance at the start from performance at the end, and dividing that value by the starting performance level (i.e. *[(b-a)/a]*). This relative change value was used as the outcome measure in a linear regression model. Predictor variables evaluated include listener group and condition. Individual characteristics (i.e. inhibition, attention, working memory, and vocabulary) were also evaluated for significant contributions to the model fit.

Generalization to a familiar accent

To examine talker-independent adaptation to the non-native accent, performance on the generalization task was compared to performance at both the starting and ending points of the adaptation period. Performance for the first and last 10 sentences within adaptation was compared to the performance during the generalization phase (10 trials). A generalized linear mixed effects regression (GLMER) model was constructed with proportion keywords correct serving as the dependent variable. A weighted binomial distribution was utilized. For this and all GLMER analyses, forward-selection model building procedures recommended by Hox and colleagues (2010) were followed. Fixed effects included: test point (start of adaptation, end of adaptation, generalization), condition, and group. Planned comparisons allowed for an evaluation of significant differences between start of adaptation and generalization, and between end of adaptation and generalization. Random effects structure included random intercepts for subject and token, and random slopes of condition by subject were evaluated for significant contribution to model fit.

Generalization to an unfamiliar accent

The analysis of accent-independent learning was modeled after that used by Baese-Berk et al. (2013). A GLMER was constructed with proportion of keywords correct as the dependent variable (including only data from the Generalization – Unfamiliar tasks). The fixed effects included Condition and Group, and the interaction of Condition and Group was evaluated for contribution to model fit. C1 served as the reference; significantly higher performance in Conditions 2-5 as compared to C1 would be interpreted to indicate that exposure to one or more non-native talkers facilitates generalization to a novel accent more than exposure to a native talker. The random effects structure included random intercepts for subject and token. Random slopes of condition by subject were evaluated for contribution to model fit.

Speech Recognition Results

Adaptation to a single talker: Effect of accent.

To assess the effect of native talker status on rapid adaptation, the data from Condition 1 (Single ENG talker) and Condition 2 (Single SPA talker) were compared. These data are shown in Figure 2.4, and the results of the GAMM analysis are presented in Table 2.4.

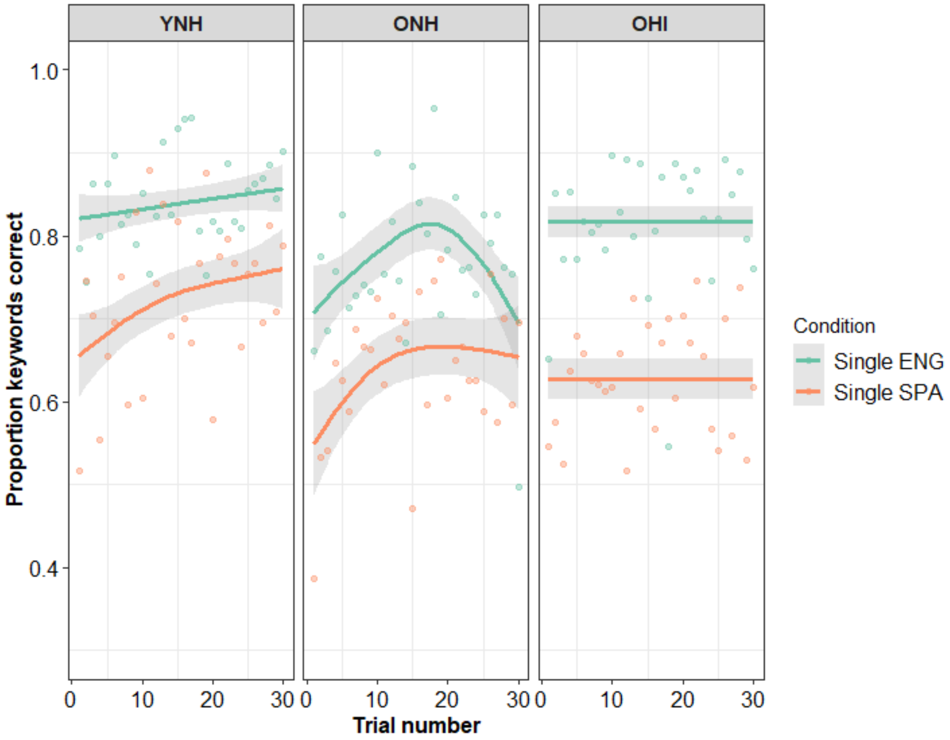


Figure 2.4. Speech recognition performance over the course of 30 trials for Conditions 1 (Single ENG; green) and 2 (Single SPA; orange), separated by listener group. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

Table 2.4. GAMM analysis comparing Conditions 1 and 2. Reference levels: Group = ONH, Condition = C1. Note that an alpha of .025 is used for significance testing due to Bonferroni correction for an ordinal-coded model.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	1.74	0.3	5.69	<.0001
Is_YNH_ordTRUE	0.67	0.29	2.28	<.025
Is_OHI_ordTRUE	0.7	0.29	2.29	<.025
Is_C2_ordTRUE	-0.81	0.33	-2.4	<.025
Is_YNH_C2_ordTRUE	-0.36	0.17	- 2.03	.04
Is_OHI_C2_ordTRUE	-0.56	0.17	- 3.08	<.01

Smooth terms	***edf***	***Ref.df***	***χ2***	***p***
s(Trial)	3.29	3.91	15.52	<.01
s(Trial): Is_YNH_ordTRUE	1.00	3.76	0.33	.57
s(Trial): Is_OHI_ ordTRUE	1.00	1.00	0.13	.72
s(Trial): Is_C2_ordTRUE	1.00	1.00	0.14	.71
s(Trial): Is_YNH_C2_ordTRUE	2.94	3.66	0.23	.63
s(Trial): Is_OHI_C2_ordTRUE	4.15	5.11	0.52	.47
s(Trial, Subject)	119.53	537.00	690.93	<.0001
s(Trial, Subject): Is_C2_ordTRUE	77.47	537.00	119.26	<.0001
s(Trial, Token)	253.03	1078.00	2374.58	<.0001

The modelling revealed that all three groups of listeners had significantly poorer performance when listening to the SPA talker than the ENG talker. This talker effect remained steady over the course of listening, and was similar across the normal-hearing groups (ONH: Parametric: *β* = -0.79, *SE* = 0.33, *z* = -2.4, *p*<.05; Smooth: edf = 1.0, χ^2 = .14, *p*=.71; ONH vs YNH Parametric: *β* = -0.34, *SE* = 0.17, *z* = -2.03, *p*=.043; Smooth: edf = 1.0, χ^2 = 0.27, *p*=.63). However, the interaction of hearing loss and condition was significant for the parametric but not smooth terms (ONH vs OHI Parametric: *β* = -0.51, *SE* = 0.16, *z* = -3.08, *p*<.01; Smooth: edf = 1.0, χ^2 = .52, *p*=.47). This indicates that the condition effect was larger for the OHI listeners than for the ONH listeners, and remained larger for the entire course of adaptation.

Further examination of this interaction using model re-leveling and re-running showed that OHI listeners had significantly higher performance for C1 than the ONH listeners, though their performance patterns did not significantly differ (Parametric: β = 0.7, *SE* = 0.3, *z* = 2.32, *p*<.025; Smooth: edf = 1.0, χ^2 = .11, *p*=.74). Similarly, ONH listeners' performance in the Single ENG condition was significantly poorer than that of the YNH listeners (β = 0.67, *SE* = 0.29, *z* = 2.28, *p*<.025). The performance pattern smooths did not differ significantly between the two normal-hearing groups (edf = 1.0, χ^2 = .33, *p*=.57). In the Single SPA condition, performance was equivalent between the two older listener groups (*p*>.025, all comparisons), as well as the two normal-hearing listener groups. In summary, when listening to a single talker, the presence of a NNE talker reduced overall speech recognition performance for all three listener groups, but did not appear to influence the pattern of rapid adaptation.

Adaptation to a single talker: Effect of L1

To assess the effect of non-native talker L1 on rapid adaptation, the data from Condition 2 (Single SPA talker) and Condition 3 (Single JPN talker) were compared. These data are shown in Figure 2.5.

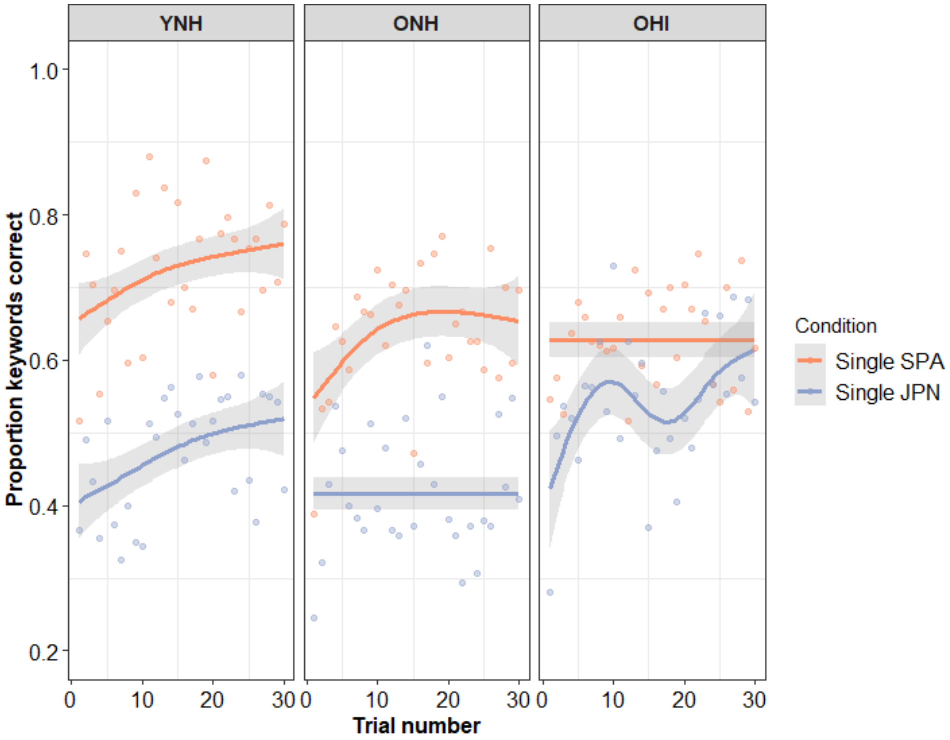


Figure 2.5. Speech recognition performance over the course of 30 trials for Conditions 2 (Single SPA; orange) and 3 (Single JPN; blue), separated by listener group. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

The GAMM analyses revealed that both of the normal-hearing listener groups had significantly poorer performance when listening to the JPN as compared to the SPA talker. This talker effect was similar for both groups, and did not change throughout the condition (Reference: Parametric: β = -1.48, *SE* = 0.31, *z* = -4.76, *p*<.001; Smooth: edf = 1.0, χ^2 = 0.2, *p*=.88; Interaction: Parametric: β = 0.08, *SE* = 0.19, *z* = 0.47, *p*=.64; Smooth: edf = 1.0, χ^2 = 0.04, *p*=.85). However, the talker effect did interact with group when compared to the OHI listeners, on the parametric but not the smooth term

(Parametric: $\beta = 0.67$, $SE = 0.19$, $z = 3.49$, $p<.001$; Smooth: edf = 4.2, $\chi^2 = 11.74$, $p=.039$). Examination of this talker × hearing loss effect showed that the talker effect was smaller for the OHI listeners as compared to the ONH listeners, though not absent for the OHI listeners (OHI Parametric: $\beta = -0.78$, $SE = 0.31$, $z = -2.47$, $p<.025$; Smooth: edf = 1, $\chi^2 = 0.3$, $p=.58$). Releveled models indicated that the ONH and OHI listeners performed similarly in C2 ($p>.025$, parametric and smooth comparisons), but that OHI listeners had higher performance than the ONH listeners in C3 (Parametric: $\beta = 0.75$, $SE = 0.26$, $z = 2.88$, $p<.01$; Smooth: edf = 3.91, $\chi^2 = 4.69$, $p=.11$). Full details of the model can be seen in Table 2.5.

Table 2.5. GAMM analysis comparing Conditions 2 and 3. Reference levels: Group = ONH, Condition = C2. Note that an alpha of .025 is used for significance testing due to Bonferroni correction for an ordinal-coded model.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	0.93	0.29	3.22	<.01
Is_YNH_ordTRUE	0.3	0.29	1.03	.31
Is_OHI_ordTRUE	0.1	0.29	0.34	.74
Is_C3_ordTRUE	-1.48	0.31	-4.76	<.0001
Is_YNH_C3_ordTRUE	0.09	0.19	0.47	.64
Is_OHI_C3_ordTRUE	0.67	0.19	3.49	<.001
Smooth terms	***edf***	***Ref.df***	***χ2***	***p***
s(Trial)	3.13	3.7	13.9	<.01
s(Trial): Is_YNH_ordTRUE	1.00	1.00	0.02	.89
s(Trial): Is_OHI_ ordTRUE	1.00	1.00	0.45	.5
s(Trial): Is_C3_ordTRUE	1.00	1.00	0.02	.88
s(Trial): Is_YNH_C3_ordTRUE	1.00	1.00	0.04	.85
s(Trial): Is_OHI_C3_ordTRUE	4.2	5.005	11.74	.04
s(Trial, Subject)	102.43	537.00	771.61	<.0001
s(Trial, Subject): Is_C3_ordTRUE	136.87	537.00	261.7	<.0001
s(Trial, Token)	300.79	1078.00	2605.05	<.0001

Overall, the comparison of these two talker types indicates that the talker’s L1 influenced listeners’ performance, with all listeners showing poorer performance with a single JPN talker as compared to a single SPA talker.

Adaptation to a single L1: Effect of multiple talkers

To assess the effect of the number of talkers on rapid adaptation, the data from Condition 2 (Single SPA talker) and Condition 4 (Multiple SPA talkers) were compared. These data are shown in Figure 2.6.

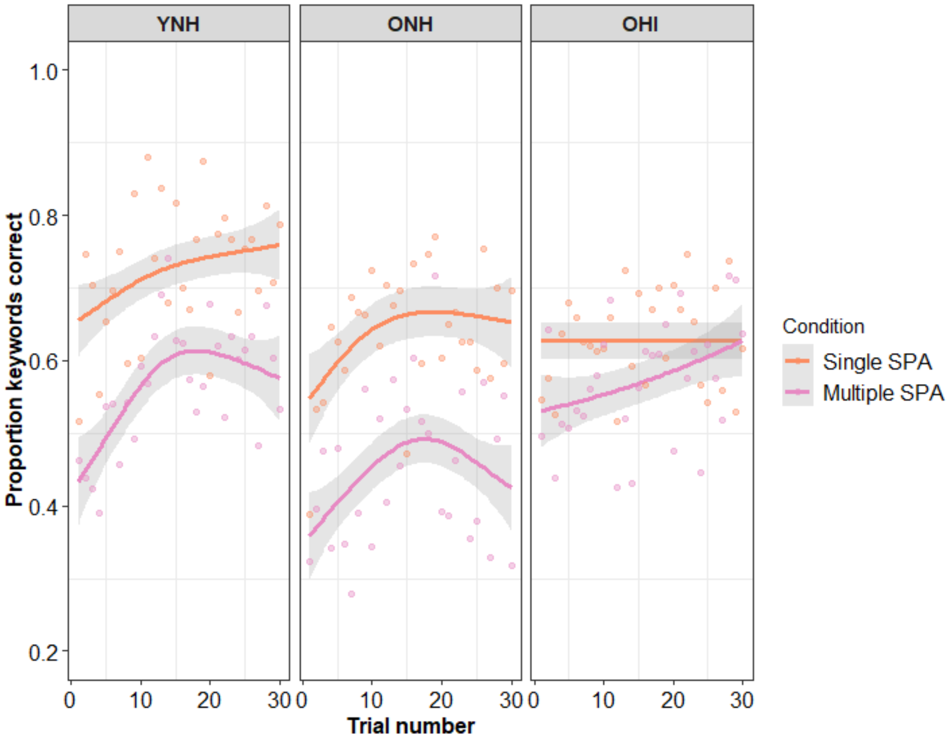


Figure 2.6. Speech recognition performance over the course of 30 trials for Conditions 2 (Single SPA; orange) and 4 (Multiple SPA; pink), separated by listener group. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

The GAMM analyses revealed that the listeners with normal hearing (both YNH and ONH) had significantly poorer performance when listening to the multiple talkers as compared to a single talker, and this effect remained steady over the course of listening (ONH Parametric: $\beta = -0.98$, $SE = 0.3$, $z =$

-2.8, *p*<.01; Smooth: edf = 1.00, χ^2 = 0.16, *p*=.69; Age × Condition Parametric: *β* = 0.31, *SE* = 0.17, *z* = 1.79, *p*=.07; Smooth: edf = 2.05, χ^2 = 2.59, *p*=.34). In contrast, the OHI listeners did not show singificanlty different performance in the single vs multiple SPA conditions (*p*>.025). When comparing the two older listener groups, there was an effect of hearing loss on the parametric but not smooth terms (HL × Condition Parametric: *β* = 0.62, *SE* = 0.17, *z* = 3.62, *p*<.001; Smooth: edf = 1.00, χ^2 = 0.38, *p*=.54), suggesting that the talker effect was smaller for the OHI listeners than for the ONH listeners, and that it remained smaller for the duration of the condition. In fact, further examination revealed that the OHI listeners' performance did not differ significantly for single vs multiple SPA talkers (*p*>.025, both parametric and smooth comparisons). Full details of the model can be found in Table 2.6.

Table 2.6. GAMM analysis comparing Conditions 2 and 4. Reference levels: Group = ONH, Condition = C2. Note that an alpha of .025 is used for significance testing due to Bonferroni correction for an ordinal-coded model.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	0.9	0.29	3.13	<.01
Is_YNH_ordTRUE	0.31	0.28	1.11	.27
Is_OHI_ordTRUE	0.13	0.28	0.46	.64
Is_C4_ordTRUE	-0.98	0.35	-2.81	<.01
Is_YNH_C4_ordTRUE	0.31	0.17	1.79	.07
Is_OHI_C4_ordTRUE	0.63	0.17	3.62	<.001
Smooth terms	***edf***	***Ref.df***	***χ2***	***p***
s(Trial)	2.72	3.21	11.91	<.01
s(Trial): Is_YNH_ordTRUE	1.00	1.00	0	.99
s(Trial): Is_OHI_ ordTRUE	1.00	1.00	0.26	.61
s(Trial): Is_C4_ordTRUE	1.00	1.00	0.16	.69
s(Trial): Is_YNH_C4_ordTRUE	2.05	2.41	2.59	.34
s(Trial): Is_OHI_C4_ordTRUE	1.00	1.00	0.38	.54
s(Trial, Subject)	133.32	537.00	845.56	<.0001
s(Trial, Subject): Is_C4_ordTRUE	108.68	537.00	189.34	<.0001
s(Trial, Token)	224.28	898.00	2374.58	<.0001

Adaptation to multiple talkers: Effect of L1 variability

To assess the effect of native talker status on rapid adaptation to multiple talkers, the data from Condition 4 (Multiple SPA talkers) and Condition 5 (Multiple L1 talkers) were compared. These data are shown in Figure 2.7.

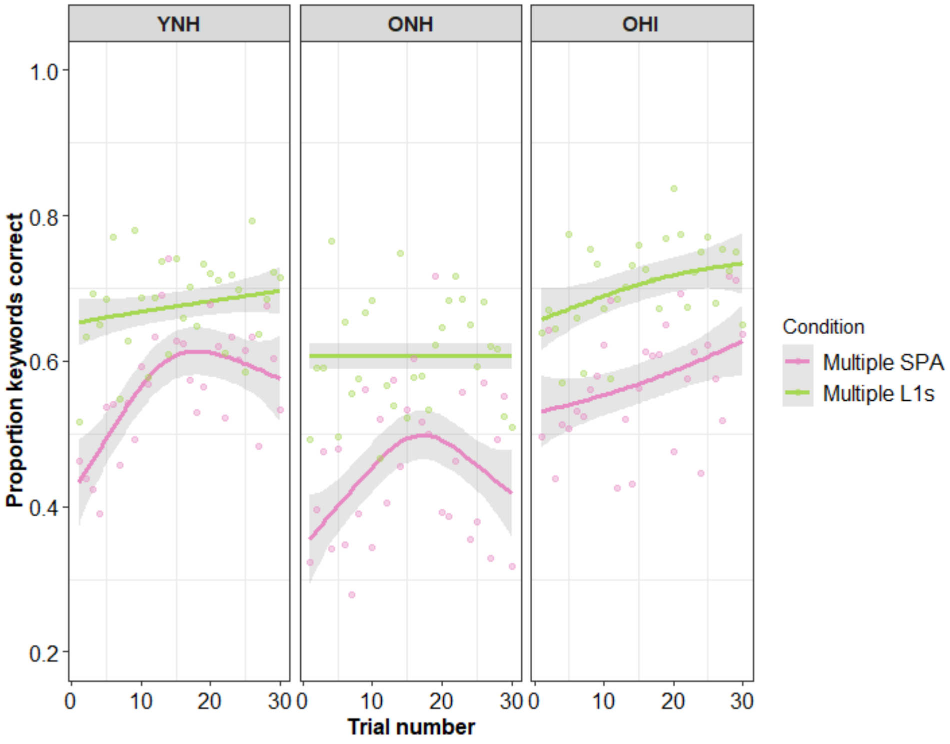


Figure 2.7. Speech recognition performance over the course of 30 trials for Conditions 4 (Multiple SPA; pink) and 5 (Multiple L1s; green), separated by listener group. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

Unexpectedly, all three listener groups had significantly poorer performance when listening to the multiple SPA talkers as compared to the multiple L1 talkers. This effect remained constant over time (Parametric: β = 0.85, SE = 0.27, z = 3.16, p<.05; Smooth: edf = 1.00, χ^2 = 0.01, p=.92). Neither of the Condition × Group interactions were found to be significant with either the parametric or smooth terms (p>.05, all comparisons), suggesting that all three listener groups showed a constant effect of L1 variability across the course of the 30 sentences, regardless of age or hearing sensitivity. Full details can be found in the model summary in Table 2.7.

Table 2.7. GAMM analysis comparing Conditions 4 and 5. Reference levels: Group = ONH, Condition = C4. Note that an alpha of .025 is used for significance testing due to Bonferroni correction for an ordinal-coded model.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	- 0.1	0.27	- 0.35	.72
Is_YNH_ordTRUE	0.63	0.25	2.52	<.25
Is_OHI_ordTRUE	0.71	0.25	2.81	<.01
Is_C5_ordTRUE	0.76	0.32	2.41	<.25
Is_YNH_C5_ordTRUE	- 0.21	0.17	- 1.26	.21
Is_OHI_C5_ordTRUE	- 0.09	0.17	- 0.50	.61
Smooth terms	***Edf***	***Ref.df***	***χ2***	***p***
s(Trial)	2.58	3.0	7.68	.05
s(Trial): Is_YNH_ordTRUE	1.89	2.2	2.07	.42
s(Trial): Is_OHI_ ordTRUE	1.00	1.00	0.09	.77
s(Trial): Is_C5_ordTRUE	1.00	1.00	0.15	.70
s(Trial): Is_YNH_C5_ordTRUE	1.61	1.88	0.92	.52
s(Trial): Is_OHI_C5_ordTRUE	1.55	1.79	0.64	.59
s(Trial, Subject)	167.07	537.00	826.59	<.0001
s(Trial, Subject): Is_C5_ordTRUE	99.47	537.00	185.88	<.0001
s(Trial, Token)	210.79	718.00	2150.4	<.0001

Magnitude of adaptation

Magnitude of adaptation was calculated by comparing performance on the first 5 and last 5 trials of the adaptation conditions. A relative change score (End-Start/Start) was calculated in order to account for differences in starting performance. A GLMER was constructed using the forward-selection model building procedures described by Hox et al. (2010) to examine the contributions of Group, Condition, and the individual predictors to predicting magnitude of adaptation. The initial model that was selected to describe the data was: $RelativeChange \sim Condition * \ Group * \ Stroop + (1|Subject)$. However, examination of the model showed that the results were clearly driven by outlier values. A subsequent model was therefore run which excluded any relative change values that were greater than 2 standard deviations above the mean relative change value. These data are plotted in Figure 2.8

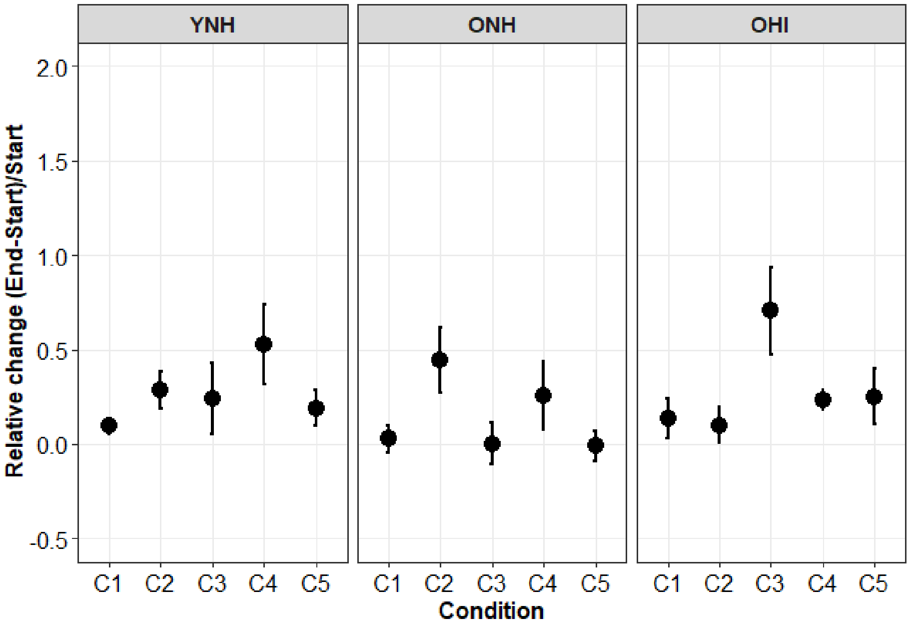


Figure 2.8. Magnitude of rapid adaptation observed in each condition, for all listeners. Outlier values falling outside of 2 standard deviations from the mean were removed from plots and analysis. Error bars reflect standard error. C1: Single ENG; C2: Single SPA; C3: Single JPN; C4: Multiple SPA; C5: Multiple language backgrounds. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

For these data, Stroop scores no longer contributed to the model fit. These data were best described by the following model: $RelativeChange \sim Condition * \ Group + (1|Subject)$. The full model summary can be found in Table 2.8.

Table 2.8. LMER for magnitude of adaptation. Reference levels: Group = ONH, Condition = C1. ICC = Intraclass correlation coefficient.

	Relative Change			
Predictors	***Estimate***	***std. Error***	***t-value***	***p***
(Intercept)	0.03	0.13	0.20	.84
Condition [C2]	0.42	0.19	2.25	<.05
Condition [C3]	-0.01	0.20	-0.05	.96
Condition [C4]	0.23	0.19	1.20	.23
Condition [C5]	-0.04	0.19	-0.21	.83
Group [YNH]	0.07	0.19	0.35	.72
Group [OHI]	0.11	0.19	0.58	.56
Condition [C2] * Group [YNH]	-0.23	0.26	-0.86	.39
Condition [C3] * Group [YNH]	0.16	0.28	0.58	.56
Condition [C4] * Group [YNH]	0.21	0.27	0.77	.44
Condition [C5] * Group [YNH]	0.14	0.27	0.52	.61
Condition [C2] * Group [OHI]	-0.46	0.26	-1.74	.08
Condition [C3] * Group [OHI]	0.58	0.27	2.13	<.05
Condition [C4] * Group [OHI]	-0.13	0.27	-0.49	.63
Condition [C5] * Group [OHI]	0.15	0.27	0.58	.56
Random Effects				
σ^2	0.35			
τ_{00} Subject	0.02			
ICC	0.05			
$N_{Subject}$	60			
Observations	290			
Marginal R^2 / Conditional R^2	0.093 / 0.136			

The significant interaction between Group and Condition was examined using the emmeans package, and showed that the source of the interaction was OHI magnitude of improvement in C3, which was significantly higher than improvement for Conditions 1 and 2 ($p<.05$, both comparisons). The magnitude of improvement for YNH and ONH listeners did not differ significantly across conditions ($p>.05$, all comparisons).

In summary, listeners globally showed a similar magnitude of adaptation across conditions, regardless of stimulus type. In the absence of other systematic condition-wise or group-wise effects, the isolated finding of increased magnitude of adaptation in C3 for OHI listeners is not thought to hold importance in understanding the variables that contribute to rapid adaptation to non-native speech.

Generalization to a familiar accent

Figure 2.9 shows the patterns of generalization to an unfamiliar talker with a familiar accent (i.e. an accent that was heard during the adaptation phase, immediately preceding generalization).

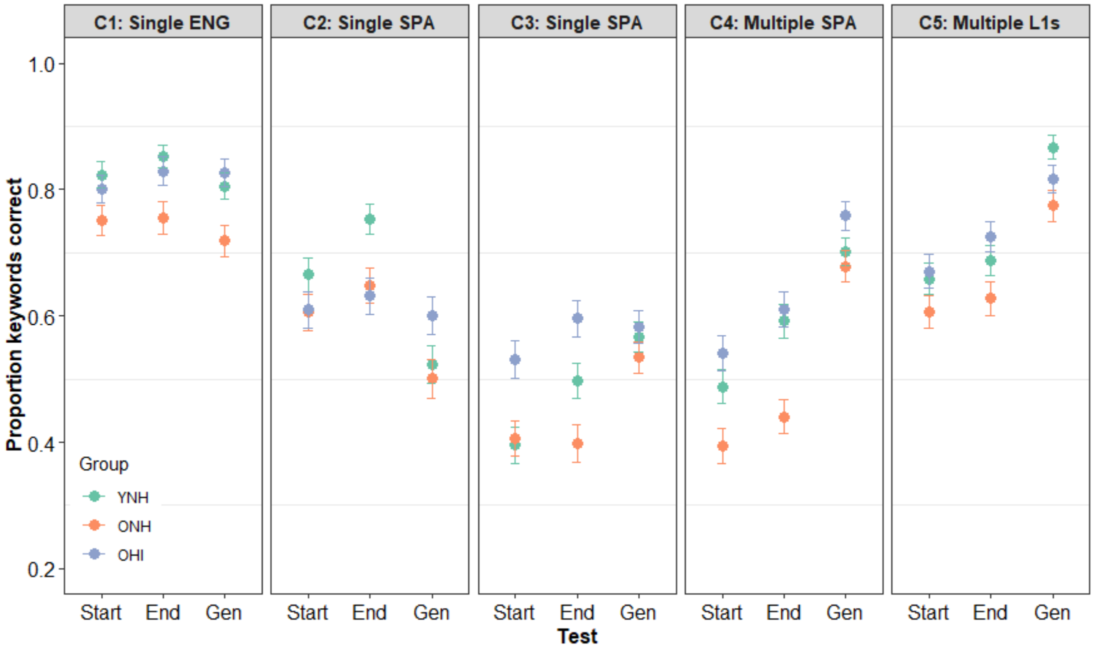


Figure 2.9. Generalization to an unfamiliar talker speaking with a familiar accent. Data displayed include the first 10 trials of adaptation (Start), the last 10 trials of adaptation (End), and the generalization phase (Gen). Error bars indicate standard error of the mean. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

In Condition 5 (C5), the generalization talker's L1 was Spanish, which is one of the multiple L1s included in C5.

The final model that was selected to describe the generalization to an unfamiliar talker with a familiar accent was a generalized linear mixed effects regression (GLMER) model with weighted binomial distribution: $PercentCorrect \sim (Test * Condition * Group) + (StroopEffect_z) + (1 + Condition|Subject) + (1|Token)$. Stroop scores were not found to interact with Group ($\chi^2(2)$= 0.99, *p*=.61) or Condition ($\chi^2(4)$= 2.75, *p*=.6). The full model output is presented in Table 2.9.

Table 2.9. GLMER for generalization to a familiar accent. Reference levels: Group = ONH, Condition = C1, Test = Generalization. ICC = Intraclass correlation coefficient

	Accuracy			
Predictors	***Odds Ratio***	***std. Error***	***z- value***	***p***
(Intercept)	3.78	.38	3.49	<.001
Condition [C2]	0.26	.48	-2.76	<.01
Condition [C3]	0.35	.48	-2.21	<.05
Condition [C4]	0.74	.48	-.63	.53
Condition [C5]	2.16	.49	1.58	.11
Test [Start]	1.43	.40	.89	.37
Test [End]	1.95	.40	1.68	.09
Group [OHI]	3.00	.30	3.72	<.001
Group [YNH]	1.57	.31	1.44	.15
StroopEffect_z	0.76	.11	-2.64	<.01
Condition [C2] * Test [Start]	1.53	.56	.76	.45
Condition [C3] * Test [Start]	0.31	.56	-2.13	<.05
Condition [C4] * Test [Start]	0.20	.57	-2.85	<.01
Condition [C5] * Test [Start]	0.17	.58	-3.05	<.01
Condition [C2] * Test [End]	1.64	.56	.88	.38
Condition [C3] * Test [End]	0.24	.56	-2.54	.011
Condition [C4] * Test [End]	0.19	.57	-2.86	<.01
Condition [C5] * Test [End]	0.17	.58	-3.09	<.01
Condition [C2] * Group [OHI]	0.51	.24	-2.86	<.01
Condition [C3] * Group [OHI]	0.46	.22	-3.50	<.001
Condition [C4] * Group [OHI]	0.60	.23	-2.28	<.05
Condition [C5] * Group [OHI]	0.35	.25	-4.13	<.001
Condition [C2] * Group [YNH]	0.91	.23	-.41	.68
Condition [C3] * Group [YNH]	0.52	.21	-3.03	<.01
Condition [C4] * Group [YNH]	0.53	.21	-2.98	<.01
Condition [C5] * Group [YNH]	0.89	.25	-.47	.64
Test [Start] * Group [OHI]	0.67	.22	-1.84	.07
Test [End] * Group [OHI]	0.66	.22	-1.85	.06
Test [Start] * Group [YNH]	0.93	.21	-.35	.73
Test [End] * Group [YNH]	0.90	.22	-.49	.62

(Condition [C2] * Test [Start]) * Group [OHI]	1.16	.29	.52	.61
(Condition [C3] * Test [Start]) *Group [OHI]	2.47	.27	3.29	<.01
(Condition [C4] * Test [Start]) *Group [OHI]	2.01	.29	2.44	<.05
(Condition [C5] * Test [Start]) * Group [OHI]	2.04	.30	2.40	<.05
(Condition [C2] * Test [End]) * Group [OHI]	1.14	.30	.43	.67
(Condition [C3] * Test [End]) * Group [OHI]	2.64	.28	3.48	<.001
(Condition [C4] * Test [End]) * Group [OHI]	1.98	.29	2.37	<.05
(Condition [C5] * Test [End]) * Group [OHI]	2.56	.30	3.12	<.01
(Condition [C2] * Test [Start]) * Group [YNH]	0.74	.29	-1.03	.30
(Condition [C3] * Test [Start]) * Group [YNH]	1.22	.27	.73	.47
(Condition [C4] * Test [Start]) * Group [YNH]	1.70	.28	1.93	.05
(Condition [C5] * Test [Start]) * Group [YNH]	0.87	.29	-.49	.62
(Condition [C2] *Test [End]) * Group [YNH]	0.88	.30	-.42	.67
(Condition [C3] * Test [End]) * Group [YNH]	1.74	.28	2.02	<.05
(Condition [C4] * Test [End]) * Group [YNH]	1.99	.28	2.46	<.05
(Condition [C5] * Test [End]) * Group [YNH]	0.68	.30	-1.27	.20
Random Effects				
σ^2	3.29			
τ_{00} token	2.05			
τ_{00} Subject	0.64			
τ_{11} Subject.ConditionC2	0.14			
τ_{11} Subject.ConditionC3	0.13			
τ_{11} Subject.ConditionC4	0.10			
τ_{11} Subject.ConditionC5	0.16			
ρ_{01} Subject.ConditionC2	-0.25			
ρ_{01} Subject.ConditionC3	-0.59			
ρ_{01} Subject.ConditionC4	-0.37			
ρ_{01} Subject.ConditionC5	-0.31			
ICC	0.44			
N_{Subject}	60			
N_{token}	360			
Observations	8990			
Marginal R^2 / Conditional R^2	0.123 / 0.511			

Examination of the three-way interaction between Condition, Test, and Group revealed the following. In Conditions 1, 2, and 3 (Single ENG, Single SPA, and Single JPN, respectively), performance at generalization is not significantly different from the start or end of adaptation for any listener group ($p>.05$, all comparisons), with the exception of the ONH listeners, who show a significant decrease at generalization as compared to the end of adaptation in Condition 2 ($\beta = -1.17$, $SE = 0.4$, $z\text{-ratio} = -2.94$, $p<.01$).

In Condition 4 (Multiple SPA), YNH listeners showed stable performance at generalization as compared to both the start ($\beta = 0.82$, $SE = 0.41$, $z\text{-ratio} = 1.99$, $p=.12$) and end ($\beta = 0.39$, $SE = 0.41$, $z\text{-ratio} = 0.95$, $p=.61$) of adaptation. However, the ONH listeners showed evidence of generalization to the unfamiliar talker with a familiar accent, with performance at generalization significantly higher than both starting performance ($\beta = 1.23$, $SE = 0.41$, $z\text{-ratio} = 3.1$, $p<.01$) and ending performance ($\beta = 0.97$, $SE = 0.41$, $z\text{-ratio} = 2.35$, $p<.05$) of adaptation. OHI listeners' performance at generalization was higher than at the start of adaptation ($\beta = 0.98$, $SD = 0.42$, $z\text{-ratio} = 2.36$, $p<.05$), but did not differ significantly between the end of adaptation and generalization ($\beta = 0.7$, $SD = 0.42$, $z\text{-ratio} = 1.67$, $p=.22$).

In Condition 5 (Multiple L1s), all three listener groups showed significantly improved performance between generalization and the start of adaptation (YNH: $\beta = 1.62$, $SE = 0.43$, $z\text{-ratio} = 3.81$, $p<.001$; ONH: $\beta = 1.41$, $SE = 0.42$, $z\text{-ratio} = 3.35$, $p<.01$; OHI: $\beta = 1.1$, $SE = 0.42$, $z\text{-ratio} = 3.35$, $p<.01$). Both normal-hearing listener groups also showed a significant improvement between the end of adaptation and generalization (YNH: $\beta = 1.61$, $SE = 0.43$, $z\text{-ratio} = 3.77$, $p<.001$; ONH: $\beta = 1.12$, $SE = 0.42$, $z\text{-ratio} =$

2.27, *p*<.05), though this same pattern was not seen for the older adults with hearing impairment (β = 0.59, *SE* = 0.43, *z-ratio* = 1.4, *p*=.34).

Inspection of the main effect of Stroop (β = -0.28, *SD* = .011, *z* = -2.64, *p*<.01) reveals that individuals with a larger Stroop effect (i.e. poorer inhibitory control) demonstrated lower speech recognition scores, at all three test conditions included in this analysis (start of adaptation, end of adaptation, and generalization to a familiar talker), in all listener groups.

In summary, listeners showed different patterns of generalization in conditions where the adaptation stimulus contained multiple vs single talkers. In the multiple talker conditions, listeners showed continued improvements in speech recognition when listening to an unfamiliar talker with a shared L1 to the talker(s) they had just heard. However, this same pattern was not seen when evaluating generalization of learning in a single-talker condition. This finding is in line with prior literature showing that prior exposure to multiple NNE talkers benefits listeners when listening to an unfamiliar NNE talker. The idea is that a degree of flexibility is required in mapping challenging acoustic input to stored lexical representations of meaning. If listeners naturally have certain category boundaries that constrain this mapping process, then the exposure to multiple talkers with provision of lexical feedback may facilitate adjustment of these internal boundaries, allowing for improved recognition for a future unfamiliar talker.

This analysis also allowed for a post-hoc comparison of the performance during the starting and ending blocks of trials, which is an analysis analogous to many prior examinations of rapid adaptation to non-native English speech (e.g. Bradlow and Bent, 2008; Wade et al., 2007). Using this metric, there were no significant differences in performance at start and end of adaptation for any listeners in the single ENG condition (*p*>.05, all groups). However, significant improvements were observed in all other conditions. YNH listeners showed significant differences between start and end of adaptation in Conditions 1-4 (*p*<.05, all comparisons), but not Condition 5 (*p*=.99). ONH listeners improved on all NNE conditions except C3 (*p>.*05). OHI listeners improved on Conditions 2 and 5, but not C3 (*p=.*5) or C4 (*p*=.08). Average performance in the first block of trials was also compared across groups, and showed that all listeners had equivalent performance during the first 10 blocks of adaptation in Conditions 1, 2, and 5. In Conditions 3 and 4, the OHI listeners had significantly higher recognition than the ONH listeners, but the two NH groups were matched for performance.

Generalization to an unfamiliar accent

Performance on sentences produced by an unfamiliar talker with an accent not heard during the course of adaptation is presented in Figure 2.10.

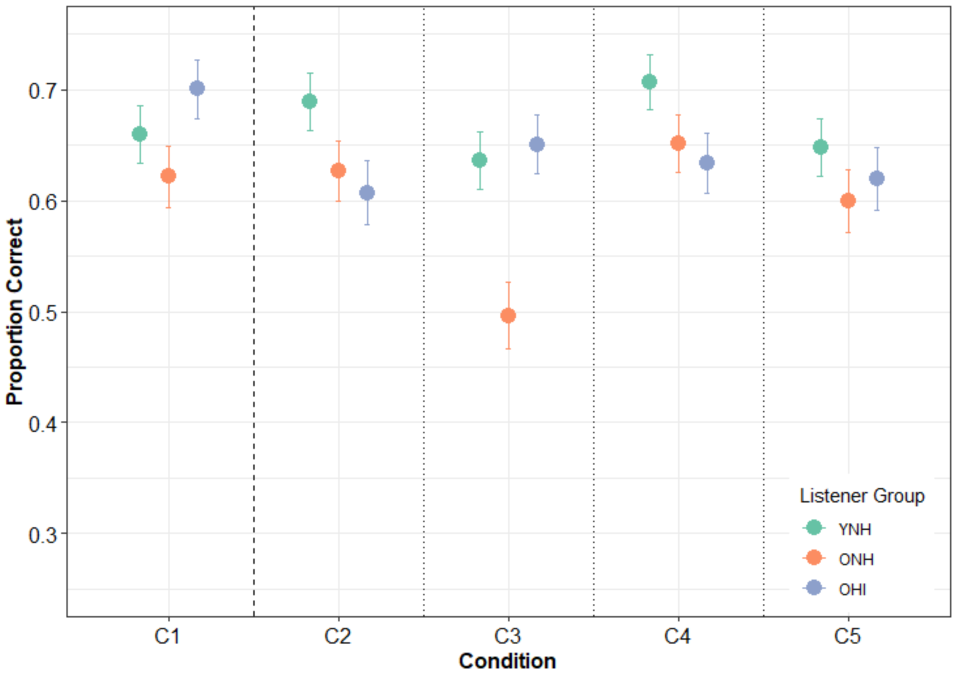


Figure 2.10. Generalization to an unfamiliar talker speaking with an unfamiliar accent. Data displayed include the generalization phase for an unfamiliar accent, separated by listener group and condition. Error bars indicate standard error of the mean. C1: Single ENG; C2: Single SPA; C3: Single JPN; C4: Multiple SPA; C5: Multiple language backgrounds. YNH: Young adults with normal hearing; ONH: Older adults with normal hearing; OHI: Older adults with hearing impairment.

These data were evaluated with respect to the effects of group and condition on speech recognition scores. If performance was higher in any of Conditions 2-5 as compared to C1, this would indicate that adaptation to non-native speech had generalized to an unfamiliar talker with an unfamiliar language background (Baese-Berk et al., 2013); differences in performance among Conditions 2-5 would indicate that exposure to certain configurations of non-native speech were more or less beneficial for generalization to an unfamiliar accent. In the model building process, neither group nor condition, nor their interaction were shown to improve model fit above a model containing random effects structure only (Group: $\chi^2(2) = 2.23$, $p=0.19$; Condition: $\chi^2(4) = 7.07$, $p=0.13$; Group × Condition: $\chi^2(14) = 22.28$, $p=0.08$). These findings suggest that exposure to non-native speech, regardless of its acoustic features, does not benefit listeners in generalizing to an unfamiliar non-native accent, above exposure to native English speech.

Individual predictors of adaptation

As performance on the Stroop test was the only individual cognitive factor to emerge as a significant predictor of speech recognition in the GLMERs described above, it was also evaluated for influence on the time course patterns of adaptation. For this analysis, Conditions 2 and 4 were selected as exemplars representing one manipulation of stimulus variability, as they compare single and multiple NNE talker conditions with talker L1 held constant. For these analyses, another GAMM was constructed utilizing combined-factor grouping variables (Van Rij et al., 2020). Stroop scores were z-scaled for analysis, and scores falling above or below 2 standard deviations of the mean were removed as outlier values. Thus, higher values of Stroop indicate a larger Stroop effect, reflecting poorer inhibitory control. Model comparison indicated that the model containing a tensor-product interaction term of Stroop significantly improved model fit (AIC difference: 34.76, $p < .05$). The summary of the full model containing the interactions with Stroop can be found in Table 2.10.

Table 2.10. GAMM examining the interactions of Stroop effect, Group, and Condition in predicting speech recognition performance as a function of trial. Reference level for parametric terms: Group = OHI, Condition = C2.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	1.2598	0.3120	4.037	<.0001
cond.allC4.OHI	-0.6417	0.3530	-1.818	.07
cond.allC2.ONH	-0.4667	0.3157	-1.478	.14
cond.allC4.ONH	-1.3207	0.4588	-2.878	<.01
cond.allC2.YNH	-0.4541	0.3995	-1.137	.26
cond.allC4.YNH	-1.0190	0.5251	-1.941	.05
Smooth terms	***Edf***	***Ref.df***	***χ2***	***p***
te(Block,StroopEffect_z):cond.allC2.OHI	3.000	3.000	9.064	<.05
te(Block,StroopEffect_z):cond.allC4.OHI	3.001	3.002	3.233	.36
te(Block,StroopEffect_z):cond.allC2.ONH	8.372	10.450	14.199	.19
te(Block,StroopEffect_z):cond.allC4.ONH	5.115	6.052	2.842	.83
te(Block,StroopEffect_z):cond.allC2.YNH	3.512	3.781	6.309	.18
te(Block,StroopEffect_z):cond.allC4.YNH	7.263	8.983	18.189	<.05
s(Trial, Subject)	145.390	498.000	1062.526	<.0001
s(Trial, Token)	239.697	898.000	2448.888	<.0001

The three-way relationship representing the interactions of Stroop effect, listener group, and condition in predicting speech recognition scores as a function of trial number are visualized using three-dimensional heatmaps. The combined-factor model displayed in Table 2.10 includes one smooth term for each combination of Stroop, group, and condition; each term can be examined in heatmap form in order to clarify the relationships between each variable in predicting speech recognition performance over trials.

An example of this heatmap visualization for the YNH listeners in Condition 2 is shown in Figure 2.11. In this figure, the x-axis represents trial number, and the y-axis represents Stroop effect, with larger values representing poorer performance on the Stroop task. The colors shown within the heatmap represent different levels of speech recognition performance, with green shades reflecting lower recognition scores, and peach shades reflecting higher speech recognition scores. The heatmap contains contour lines printed in black that serve to show how each level of speech recognition performance is predicted as a function of Stroop performance and trial number; i.e., at what values of Stroop and trial number is speech recognition performance level the same? Heatmap contour lines are marked with numbers indicating the specific speech recognition level. For example, tracking the ‘0.7’ line shows that a recognition score of 0.7 is achieved by trial number 5 for a listener with a Stroop effect of -0.5, but not until trial number 25 for a listener with a Stroop effect of 0.

To aid in comprehension of the three-dimensional heatmaps, two-dimensional ‘slices’ can be taken from the heatmap and plotted; in Figure 2.11, three such slices are plotted on the right. The white dashed lines on the heatmap represent the values of Stroop at which the slices were taken: Stroop = -1, 0, and 1. In the individual plots on the right, predicted speech recognition performance is plotted as a function of trial number for these given levels of Stroop performance. These comparisons allow for observation of how the individual’s capacity for inhibitory control influences their patterns of speech recognition performance over trials, and at which values of Stroop and trial number speech recognition performance is similar. For example, the listener with a Stroop effect of -1 (better inhibitory control) starts at a performance level of approximately 70% correct at the first trial, while the listener with a Stroop effect of 0 (mean value) only achieves that level of speech recognition performance after 25 trials. These visualizations also indicate that, in Condition 2, YNH listeners who have smaller Stroop effects (i.e. better inhibitory control) are predicted to show consistently higher performance across trials than those who have a larger Stroop effect (i.e. poorer inhibitory control).

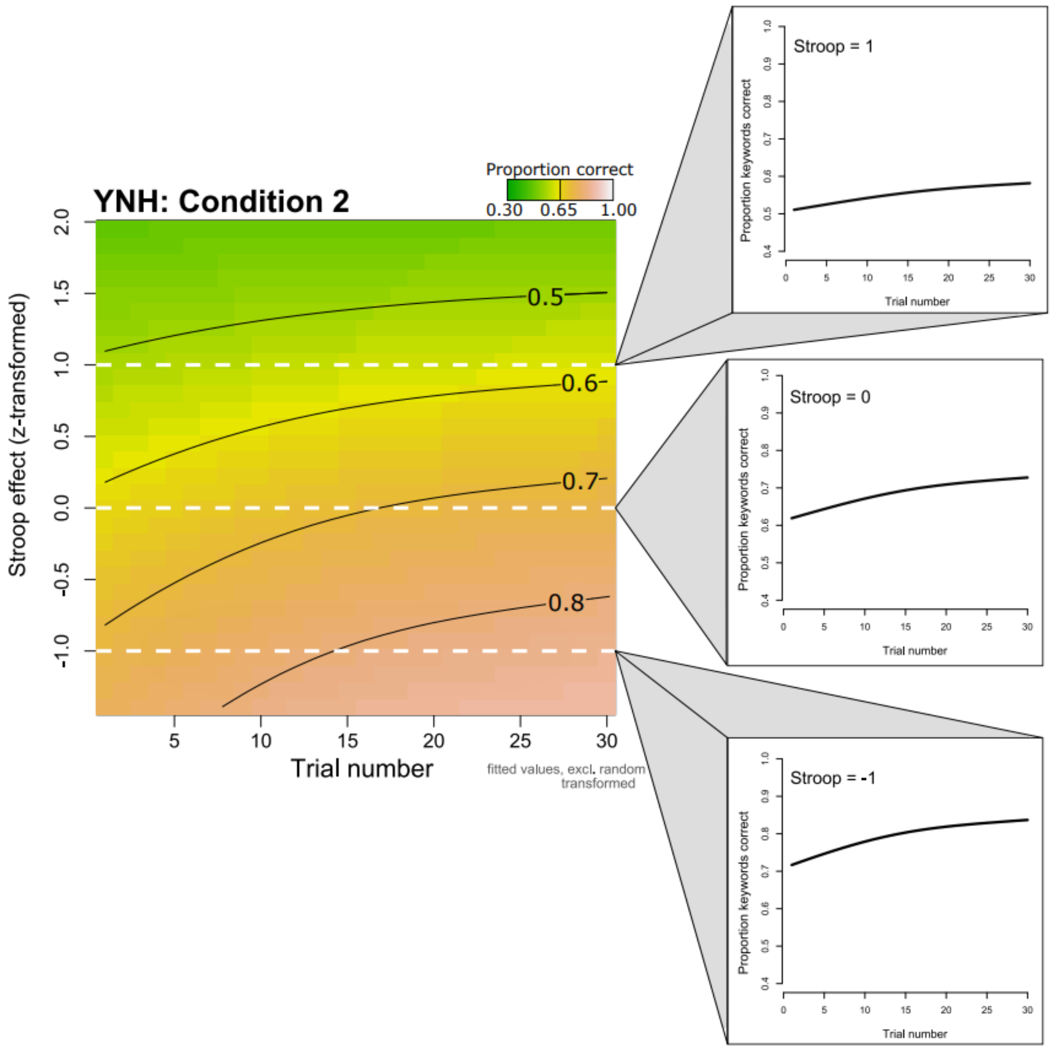


Figure 2.11. Left: a visualization of the interaction of Stroop effect (z-transformed, y-axis) and trial number (x-axis) on speech recognition performance for young normal-hearing listeners in Condition 2 (Single SPA). Shades of color represent different levels of speech recognition performance. Contour lines track levels of speech recognition performance; numbers within the contour lines indicate proportion correct. Three ‘slices’ of the heatmap (white dashed lines) are displayed on the right, representing predicted speech recognition performance over trials for individuals with Stroop effect scores of 1 (top), 0 (middle), and -1 (bottom).

Full visualizations of the relationship between Stroop performance and speech recognition as a function of trial number are shown for all listener groups along with explanatory slice plots, and are separated for Conditions 2 and 4 in Figures 2.12-2.14.

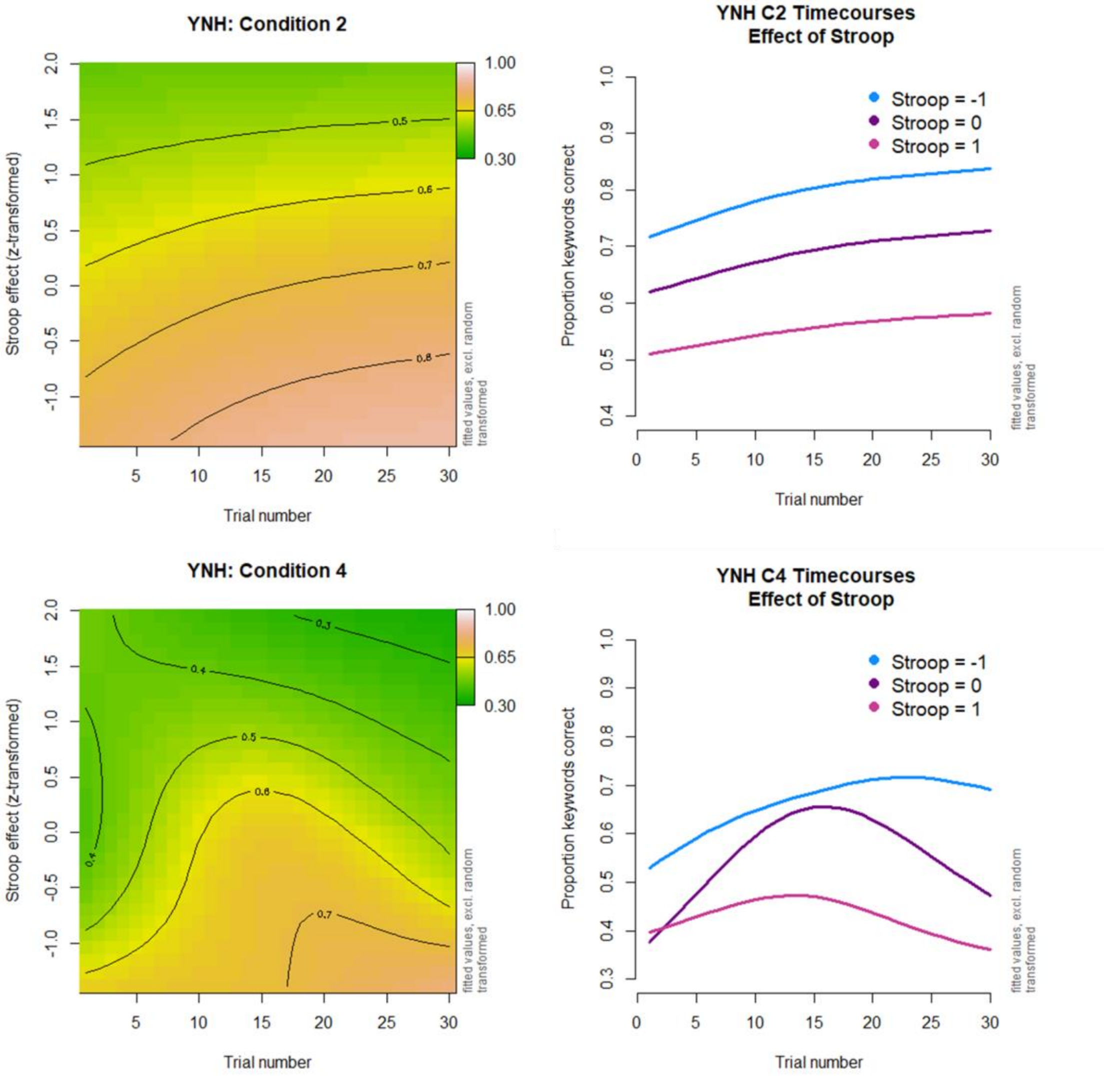


Figure 2.12. Visualizations of the interaction of Stroop effect and trial number on speech recognition performance for young normal-hearing listeners in Conditions 2 (Single SPA; top; re-plotted from Fig 2.11) and 4 (Multiple SPA; bottom). Shades of color in the heatmaps (left figures) represent levels of speech recognition. On the right side, three ‘slices’ are shown, representing predicted performance patterns for individuals with different sizes of the Stroop effect.

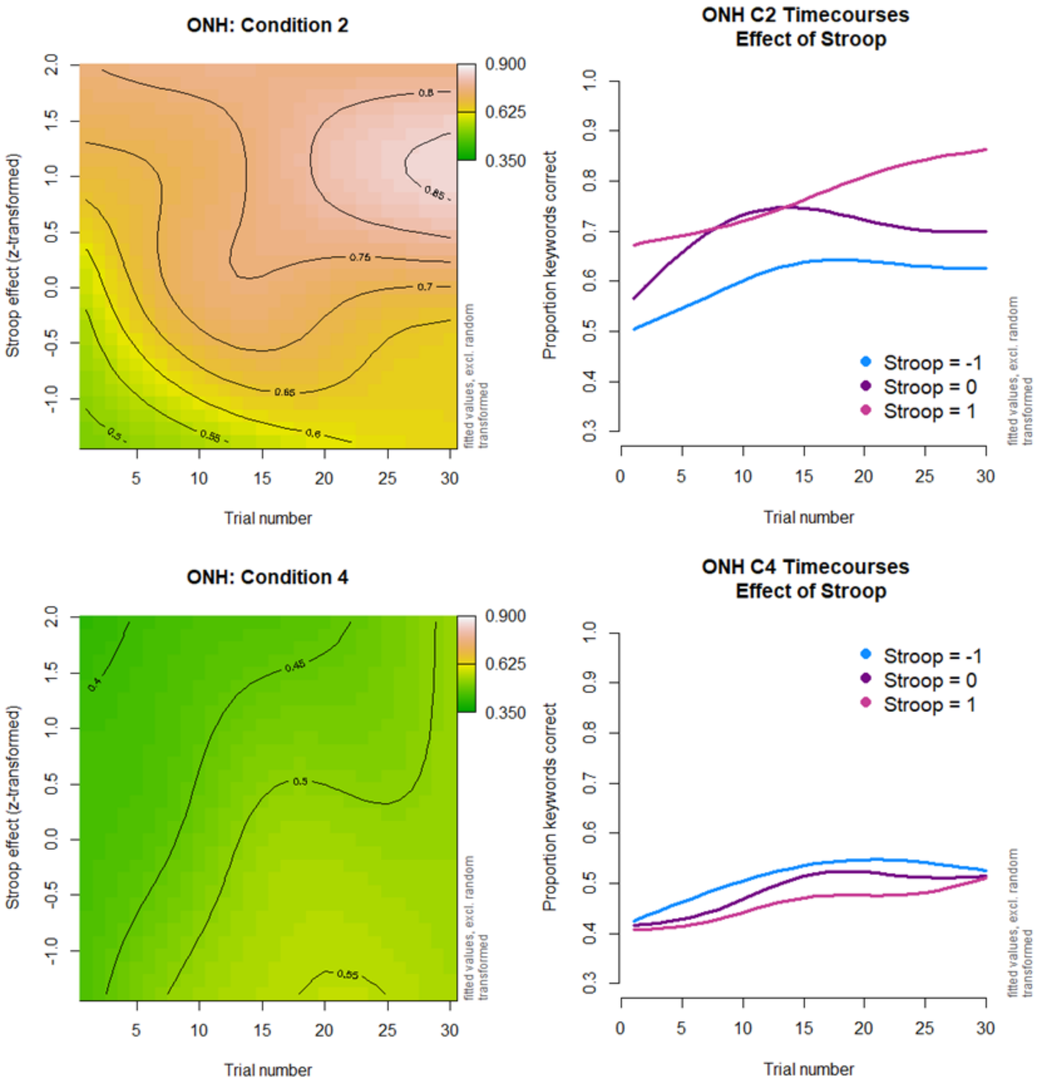


Figure 2.13. Visualizations of the interaction of Stroop effect and trial number on speech recognition performance for older normal-hearing listeners in Conditions 2 (Single SPA; top) and 4 (Multiple SPA; bottom). Shades of color in the heatmaps on the left panels represent levels of speech recognition. On the right side, three ‘slices’ are shown, representing predicted performance patterns for individuals with different sizes of the Stroop effect.

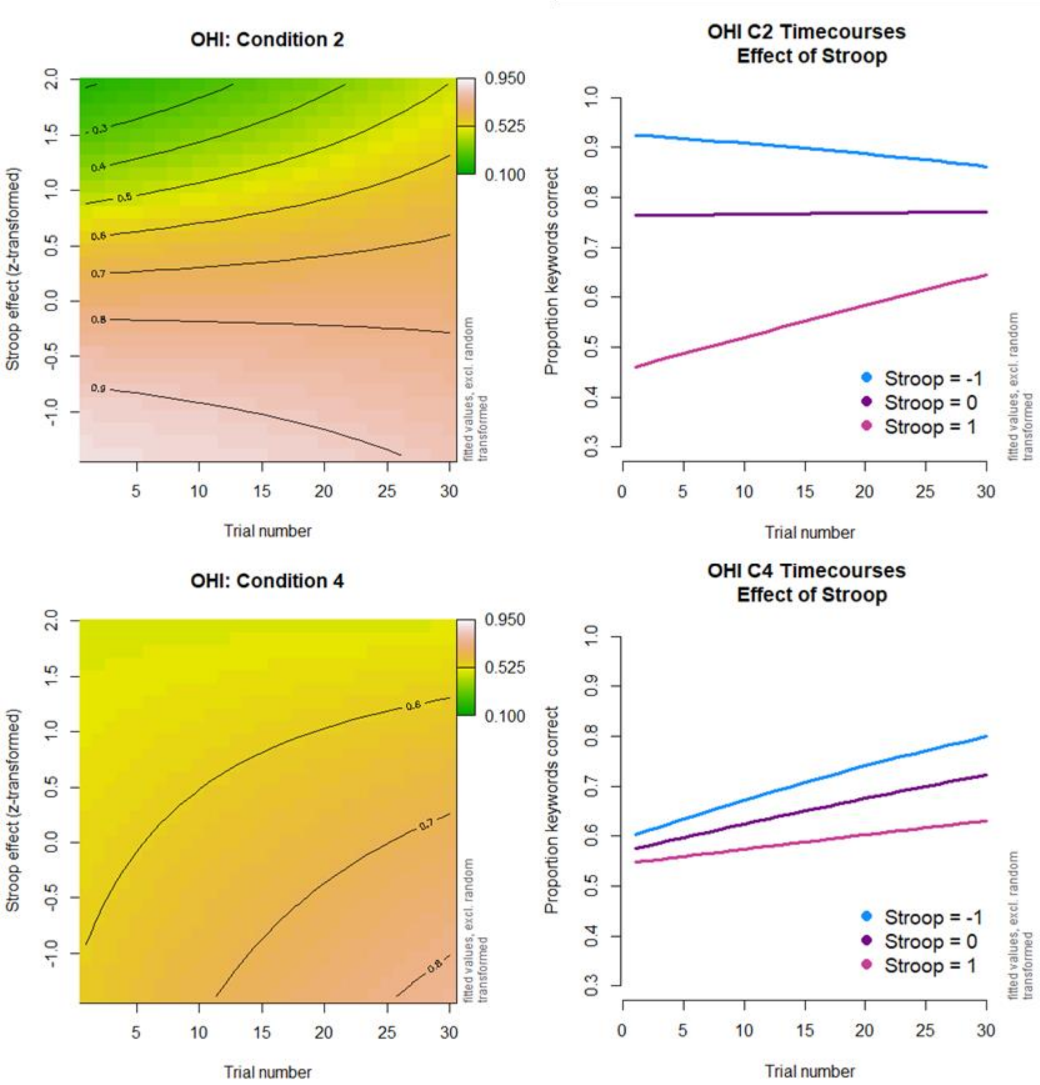


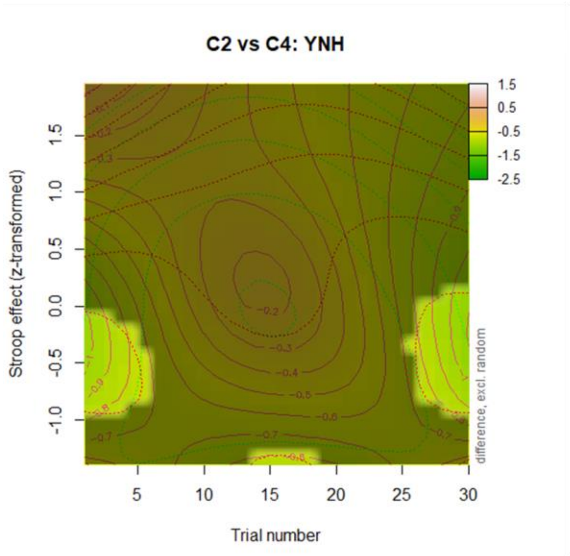
Figure 2.14. Visualizations of the interaction of Stroop effect and trial number on speech recognition performance for older hearing-impaired listeners in Conditions 2 (Single SPA; top) and 4 (Multiple SPA; bottom). Shades of color in the heatmaps on the left panels represent levels of speech recognition. On the right side, three ‘slices’ are shown, representing predicted performance patterns for individuals with different sizes of the Stroop effect.

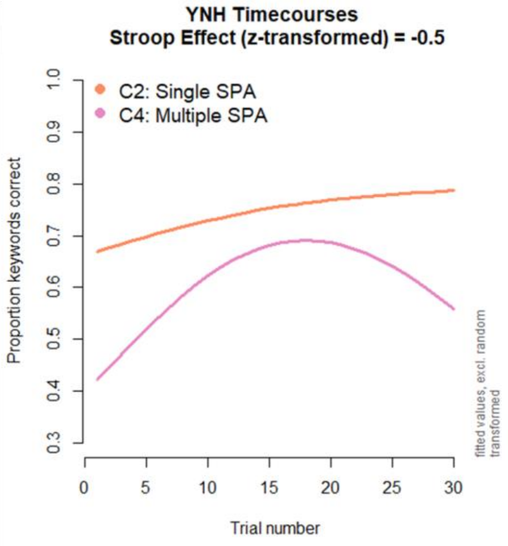
With the exception of ONH performance in Condition 2, the heatmaps generally indicate that greater Stroop effect values were associated with poorer speech recognition performance, and lower Stroop values were associated with higher speech recognition scores. In some cases, Stroop performance also appears to predict the shape of the speech recognition performance over trials. For example, in the top right panel of Figure 2.12, the patterns of adaptation are similar at the three levels of Stroop score shown for YNH listeners in C2. Examination of the heatmap for YNH listeners' performance in Condition 4 indicates that the patterns differ for different levels of Stroop performance. For example, the predicted C4 trajectory for a YNH listener with a normalized Stroop score of 0 would include an initial increase in performance (shown by the transition from green to yellow to peach colors on the heatmap), followed by a decline (transition from peach to yellow to green colors). This predicted pattern is plotted in the bottom right panel of Figure 2.12, along with the patterns for normalized Stroop scores of -1 and 1, for reference.

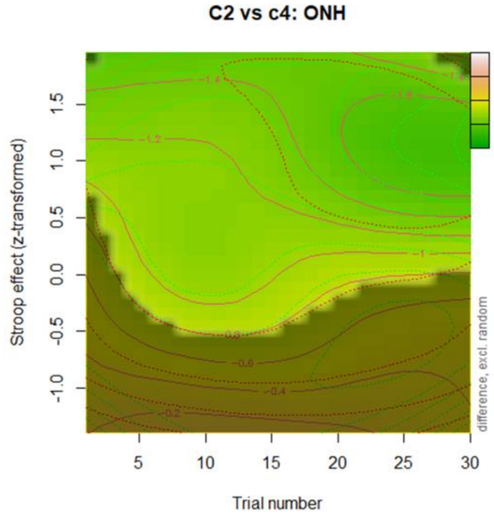
This analysis also allows for comparison of the effect of Stroop performance across conditions and groups. Significant differences would indicate that an individual's capacity for inhibitory control predicts speech recognition performance patterns differently depending on the listening conditions. In order to visualize these effects, *difference* heatmaps were generated.

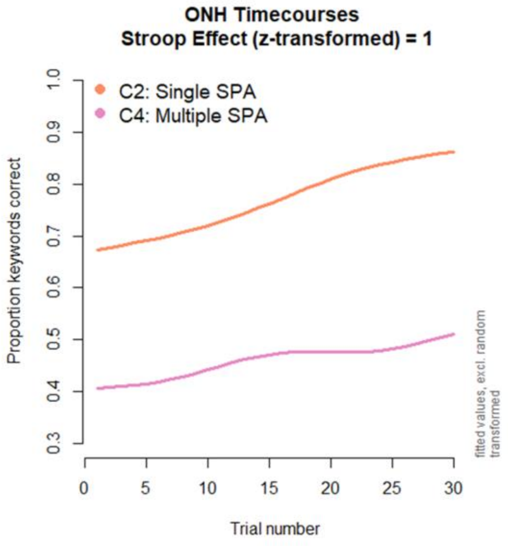
First, the condition effects were examined for each group. The difference heatmaps for each group, along with predicted adaptation patterns in the regions of significant difference, are shown in Figure 2.15. In the difference heatmaps, the x and y-axes remain the same, reflecting trial number and Stroop effect, respectively. However, the colors represent the difference in speech recognition performance between the two conditions, rather than the absolute level of performance. That is, the green shades reflect regions where performance is higher in Condition 2 (Single SPA) as compared to Condition 4 (Multiple SPA), and the peach shades reflect regions where performance is higher in Condition 4. In these figures, the highlighted areas of bright green and yellow represent the regions of significant difference; darkened areas are those where the relationships between trial number, Stroop effect, and speech recognition in the two conditions are statistically similar.

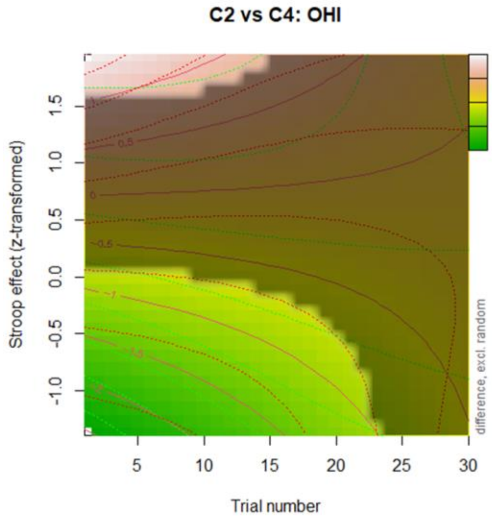
For the YNH listeners, the regions of difference are significant for listeners with Stroop effects that were smaller than the mean (i.e., better Stroop performance); listeners with larger Stroop effects (poorer inhibitory mechanisms) had statistically similar time-course patterns. Thus, young adult listeners with better Stroop performance had different trajectories across the two conditions (Conditions 2 and 4). The predicted adaptation patterns for a YNH listener with a normalized Stroop score of -0.5 are displayed in the top right panel of Figure 2.15, illustrating that performance differed between conditions at the start and end of the adaptation trials (shown on the heatmap as bright green areas in early and late trials between Stroop scores of 0 and -1), but that these listeners had similar performance in the middle of the condition.


C2 vs C4: YNH
Stroop effect (z-transformed)
Trial number
difference, excl. random
1.5
0.5
-0.5
-1.5
-2.5
-1.0
-0.5
0.0
0.5
1.0
1.5
5
10
15
20
25
30


YNH Timecourses
Stroop Effect (z-transformed) = -0.5
C2: Single SPA
C4: Multiple SPA
Proportion keywords correct
Trial number
fitted values, excl. random
transformed
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30


C2 vs c4: ONH
Stroop effect (z-transformed)
Trial number
difference, excl. random
1.5
0.5
-0.5
-1.5
-2.5
-1.0
-0.5
0.0
0.5
1.0
1.5
5
10
15
20
25
30


ONH Timecourses
Stroop Effect (z-transformed) = 1
C2: Single SPA
C4: Multiple SPA
Proportion keywords correct
Trial number
fitted values, excl. random
transformed
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30


C2 vs C4: OHI
Stroop effect (z-transformed)
Trial number
difference, excl. random
1.5
0.5
-0.5
-1.5
-2.5
-1.0
-0.5
0.0
0.5
1.0
1.5
5
10
15
20
25
30

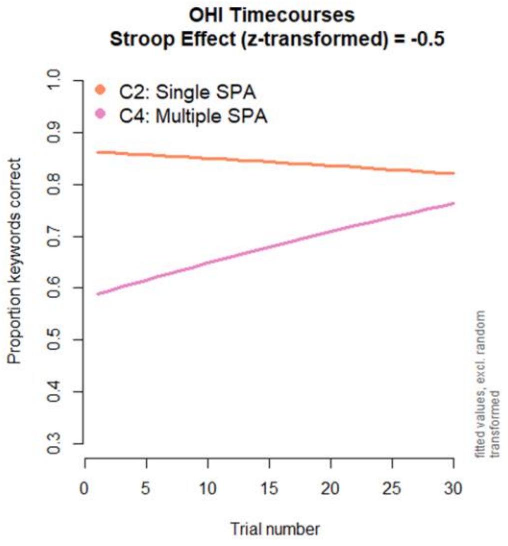

OHI Timecourses
Stroop Effect (z-transformed) = -0.5
C2: Single SPA
C4: Multiple SPA
Proportion keywords correct
Trial number
fitted values, excl. random
transformed
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30

Figure 2.15. Differences in the effect of Stroop on adaptation patterns in Conditions 2 (Single ENG) and 4 (Single SPA) for YNH (top), ONH (middle) and OHI (bottom) listeners. In the left panels, colors represent the predicted difference in speech recognition scores between the two conditions, in log-odds. Highlighted areas (i.e., regions with bright colors) indicate regions of significant difference; darkened areas are those where the relationships between trial number, Stroop effect, and speech recognition in the two conditions are statistically similar. Right-side panels show ‘slices’ with the predicted adaptation patterns for Conditions 2 and 4 for listeners with the same Stroop effect score.

The difference in Stroop effect between conditions for ONH listeners is displayed in the center panels of Figure 2.15. Here, there is a broad region of significant difference between the two conditions for listeners with larger magnitude Stroop effects (poorer inhibitory mechanisms) indicated by the highlighted areas. This condition-wise difference is larger in the later portion of trials, as indicated by the differing shades of green within the highlighted region when comparing the left and right-hand sides. This difference is illustrated in a two-dimensional view in the center right panel of Figure 2.15, where predicted performance in both conditions is shown for a listener with a Stroop effect falling one standard deviation above the mean. Overall, for the ONH listeners, predicted performance for Condition 4 is significantly lower than for Condition 2, and is consistently poorer across trials for listeners with larger magnitude Stroop effects.

The comparison of conditions for OHI listeners is visualized in the bottom panels of Figure 2.15. For these listeners, the region of significant difference occurred in the start of trials, for listeners with smaller magnitude

Stroop effects. For those listeners, the predicted performance patterns indicate that there were differences in speech recognition performance at the start of trials, but that listeners' performance in Condition 4 improved such that there were no differences in performance between conditions by the end of the sentence set. The effect of condition is also larger for individuals with smaller magnitude Stroop effects (note the highlighted region of significant difference ends around trial number 7 for a Stroop value of 0, but persists out past trial 20 for listeners with a Stroop effect of -1).

Differences between groups in each condition were also examined, though it should be noted that these group-wise comparisons are limited as the distributions of Stroop effect between younger and older normal-hearing listener groups did differ (see Table 2.1). When comparing the two normal-hearing groups, there were no significant differences in either condition; this null result should be interpreted with caution in light of the distribution differences. However, the comparison of ONH and OHI did result in significant differences in both Conditions 2 and 4. These difference heatmaps between ONH and OHI listeners, created separately for Conditions 2 and 4, are visualized in Figure 2.16. Brighter green shading reflects better performance for the ONH listeners than the OHI listeners, peach shading reflects better performance for the OHI listeners than the ONH listeners, and dark regions reflect no significant difference in performance between the two listener groups.

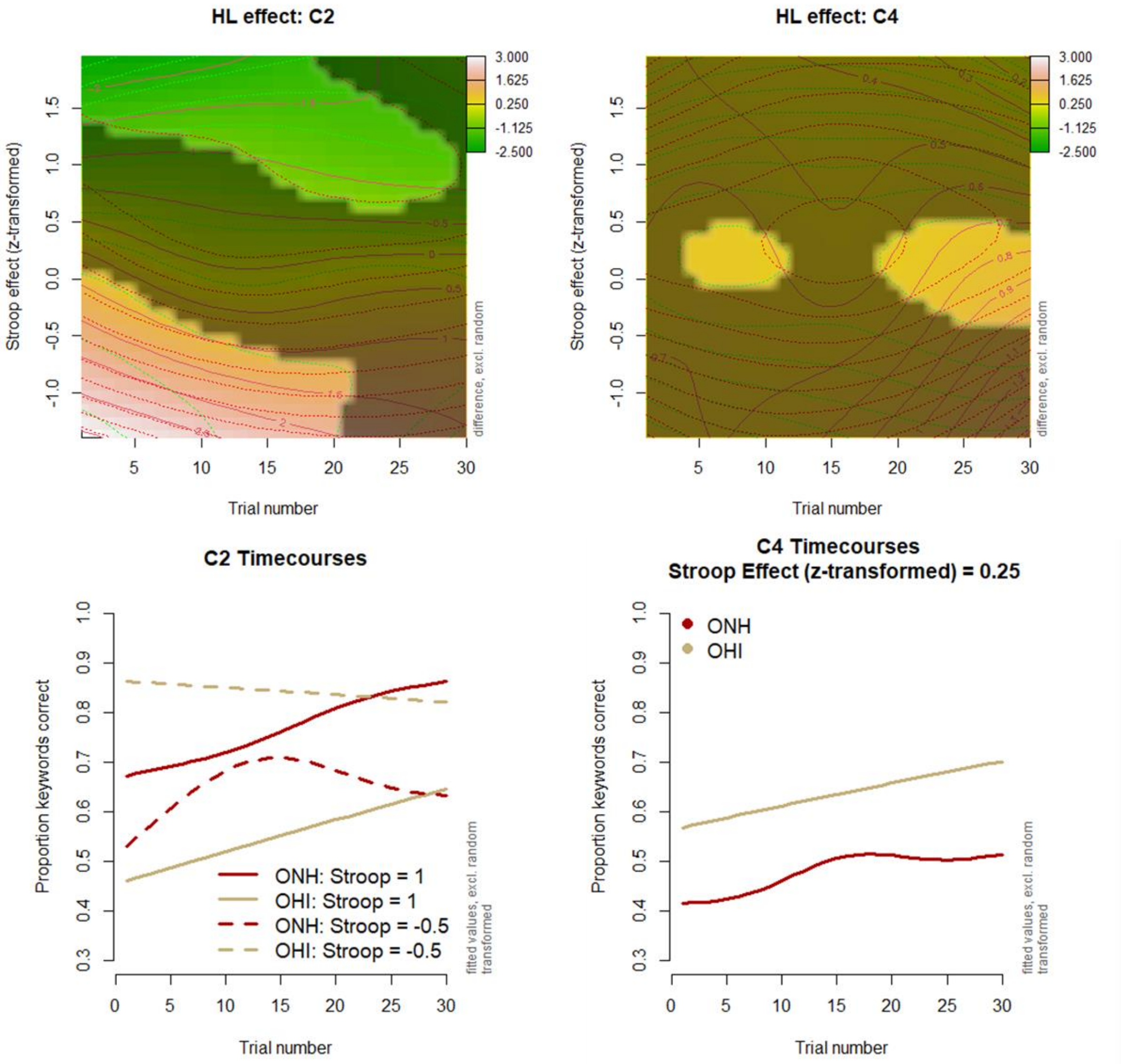


Figure 2.16. Differences in the effect of Stroop on adaptation patterns in Conditions 2 (left; Single ENG) and 4 (right; Multiple SPA) between the two older listener groups. The top row shows difference heatmaps: colors represent the predicted difference in speech recognition scores between the two groups. Highlighted areas indicate regions of significant difference; darkened areas are those where the relationships between trial number, Stroop effect, and speech recognition between the two groups are statistically similar. In the bottom row, the predicted adaptation patterns for ONH and OHI listeners with the same Stroop effect score.

In Condition 2 (Single SPA), there were two regions of significant difference, both occurring during the earlier portion of trials, and occurring in opposite directions. For listeners with Stroop effects larger than the mean (i.e. poorer inhibitory control), hearing impairment reduced speech recognition performance in Condition 2. For those with smaller Stroop effects (i.e. better inhibitory control), OHI listeners actually had higher performance than ONH, at least at the start of adaptation. These differences appear to be driven by a reversal of the expected effect of Stroop for the ONH listeners in Condition 2. For the OHI listeners, a larger magnitude Stroop effect was associated with poorer speech recognition scores, and vice versa. This pattern was also seen for overall speech recognition scores in the generalization analysis. However, the ONH listeners appeared to show the opposite effect in Condition 2, with larger magnitude Stroop effects associated with higher predicted speech recognition scores.

The group effect for Condition 4 was isolated to a smaller range of Stroop effect scores, centered around the mean. Here, when listeners had Stroop scores in the mean range, the ONH listeners performed more poorly than the OHI listeners at the start and end of adaptation, though they showed equivalent performance in the middle portion of trials. For listeners outside this mean range of Stroop effect scores, the time course patterns were similar across conditions.

Discussion

The results of this study indicate that when listeners are presented with an unfamiliar talker with a non-native accent, their speech recognition performance changes over the course of listening to 30 sentences.

Improvements in performance are observed for most groups and conditions, but these improvements were dependent on the listener group and type of stimulus.

Effects of aging on rapid adaptation to NNE speech.

In prior work, comparisons of younger and older adults with normal hearing sensitivity have indicated that aging is associated with poorer non-native speech recognition (Burda et al., 2003; Gordon-Salant et al., 2010a, 2010b), but the limited prior investigations of adaptation to NNE speech in younger and older NH listeners suggest that aging alone does not lead to reductions in the rate or magnitude of rapid adaptation to NNE speech (Adank & Janse, 2010; Gordon-Salant, Yeni-Komshian, Fitzgibbons, et al., 2010). The results of this study suggest that aging did not impair recognition or adaptation when listening to NNE speech. In all NNE talker conditions, with the exception of Condition 5 (Multiple L1s), the older and younger NH adults did not significantly differ in their recognition of the NNE speech. Additionally, when the NNE talker conditions were compared, both younger and older adults showed the same effects; the changes in talker type affected the overall level of performance, but not the time course of adaptation. Magnitude of adaptation was also found to be similar in younger and older NH listeners, across all listening conditions. Overall, aging did not appear to have a significant impact on rate or magnitude of rapid adaptation to non-native speech.

Effects of hearing loss on rapid adaptation to NNE speech

In this study, two groups of older listeners were compared for differences in rapid adaptation performance. There is very limited prior work examining the specific effects of hearing loss on rapid adaptation to non-native speech, beyond the effects of aging alone. The results of this study showed that, for most conditions and comparisons, the two older listener groups did not perform similarly. Contrary to expectations, OHI listeners actually showed higher overall performance in most conditions as compared to ONH listeners.

There were also differences in the size of the talker effects between the ONH and OHI listeners. When the single ENG and single SPA conditions were compared, the talker effect was larger for the OHI listeners than the ONH listeners. In contrast, when comparing different L1s of single NNE talkers (C2 vs C3) and single vs multiple NNE (C2 vs C4) talkers, the OHI listeners showed reduced or absent effects of talker type. When examining the magnitude of adaptation, the OHI listeners performed similarly to the ONH listeners, with the exception of the Single JPN condition, where OHI listeners showed a greater magnitude of adaptation than ONH listeners. OHI listeners also showed generalization of learning to an unfamiliar talker in the multiple talker conditions, similar to the ONH listeners.

The direct comparison between ONH and OHI listeners' performance in this study is complicated by the use of different SNRs between the two older listener groups during testing. These SNRs were designed to target similar starting levels of performance across listener groups, and in fact *post-hoc* analyses comparing the average performance on the first 10 trials of adaptation showed no significant differences across the listener groups in Conditions 1, 2, and 5, and only an ONH/OHI difference in Condition 4. These differences in starting level across condition might suggest that the effect of multi-talker babble on speech recognition performance differs depending on the characteristics of the target speech. Peelle and Wingfield (2005) found that YNH and ONH listeners did not show different patterns of adaptation to noise-vocoded speech when matched for starting performance, but did differ when matched for level of signal distortion. It is possible that the matching procedure used in this study masked some group-wise differences in pattern of adaptation. Future investigations may benefit from utilizing individually adapted SNRs, however this strategy is challenging to implement in studies of rapid adaptation, as the listener is necessarily exposed to the target stimulus during the SNR-setting procedure, and thus some adaptation may occur prior to the onset of the experimental conditions.

Effects of talker type on recognition and rapid adaptation

Five listening conditions were evaluated for rapid adaptation in this study: Single ENG, Single SPA, Single JPN, Multiple SPA, and Multiple L1s. These conditions involved different levels of indexical variability as defined by numbers of talkers and language backgrounds. Acoustic analyses indicated that the non-native English sentences were longer and more variable in duration and typically slower in rate than the native English sentences, consistent with prior reports of non-native English speech (Guion et al., 2000). A sampling of the talkers' vowel productions indicated that the non-native conditions had smaller separability between the confusable vowel pair /i/ and /ɪ/ than the native ENG talker condition, though the variability of the formants for these two vowels was higher in some but not all of the NNE conditions as compared to the NE talker condition. Overall, the acoustic analyses confirmed that there was more variability in certain acoustic measures in the non-native English speaker conditions compared to the single NE speaker condition.

Speech recognition scores were examined over the course of trials using a generalized additive mixed model (GAMM) analysis, which allowed for distinctions between intercept-level and slope-level differences in condition. These analyses overwhelmingly indicated that changes in the number and type of talkers resulted in intercept-level differences, as did the comparison between single talkers from different language backgrounds (Single SPA vs Single JPN). For the normal-hearing listener groups, the comparison of different talker conditions revealed reductions in speech recognition performance. The OHI listeners showed smaller talker effects for two of the comparisons as compared to the ONH listeners, but showed similar or exaggerated effects for the other two conditions, as described above. However, these same condition comparisons did not reveal any statistically significant changes in the patterns of adaptation. Additionally, the magnitude of adaptation was similar across nearly all conditions. Together, these findings indicate that, at least for the normal hearing listeners, increases in stimulus variability did not reduce the rate or magnitude of rapid adaptation to non-native speech, despite reducing overall speech recognition levels. While the two single NNE talker conditions were grossly similar in their variability measures, the lower performance with Japanese-accented English as compared to Spanish-accented English may relate to the relatively higher prevalence of native Spanish speakers in the Maryland/DC/Virginia area.

A small number of prior studies have examined the effects of stimulus variability specifically on the rate of rapid adaptation to NNE speech (Bradlow & Bent, 2008; Luthra et al., 2021; Wade et al., 2007; Witteman et al., 2014). These studies have generally observed that increased stimulus variability leads to a slower rate of rapid adaptation, using variability manipulations including single vs multiple talkers, single vs multiple accents, and artificial manipulations of acoustic features. In these prior studies, rate of learning was measured through methods comparing averaged blocks of trials. The present study is unique in using a nonlinear curve-fitting modelling analysis, which indicated that the overall patterns of adaptation did not differ across conditions that differed by number and/or type of talkers. However, an analysis analogous to those used in prior studies was conducted when examining generalization to new talkers. In this analysis, the first and last 10 trials of each condition were compared, which did show some differences across conditions and listener groups.

A comparison of two averaged blocks provides a different picture of listener performance than a trial-by-trial performance curve, though both types of analyses can provide important information about how listeners perform when exposed to unfamiliar or challenging speech stimuli. A comparison of blocks may be more relevant for questions about overall improvement, while detailed time course analyses can give more detail about how the listeners achieve this overall level of change. Availability of these different methods can allow researchers to utilize analyses that reflect their specific research questions when investigating rapid adaptation to challenging speech.

Generalization of learning

The results of this study indicate that patterns of generalization differ based on the stimuli that were heard during adaptation. Following a session of listening to several non-native talkers, younger and older adults were able to generalize their learning to an unfamiliar talker with a familiar language background, demonstrating significantly improved performance at generalization as compared to the start of adaptation. These same improvements were not seen for conditions in which the listeners adapted to a single talker. This finding is consistent with prior literature documenting a benefit of stimulus variability in facilitating generalization of learning (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Lively et al., 1993; Sidaras et al., 2009).

During the period of rapid adaptation, listeners are thought to use information from the target speech stream to flexibly update their internal representations of lexical information. The process of converting acoustic information into a meaningful semantic message requires that the speech stream be divided into separate phonetic units, which can be matched with the listeners' internal categories. If a listener has narrow or inflexible boundaries for these categories, then perception of unfamiliar, challenging speech should be impeded. The literature suggests that exposure to a relatively greater degree of variability during auditory perceptual learning helps listeners develop and maintain more flexible internal category boundaries, which benefits listeners when listening to an unfamiliar talker.

The results of this study support these concepts, to a certain degree. When listeners' generalization was tested on performance with unfamiliar talkers, they showed generalization to unfamiliar talkers who shared language backgrounds with the talkers they had just heard during adaptation. However, when the generalization talkers had both unfamiliar voices and unfamiliar language backgrounds, there was no evidence of any generalization. This finding, particularly the lack of generalization to an unfamiliar talker following exposure to talkers from multiple language backgrounds (C5), contrasts with prior findings from Baese-Berk et al (2013).

Xie and Myers (2017) further examined the notion that talker variability facilitates generalization to new talkers in an online lexical decision task. They found that generalization was most strongly facilitated in conditions where the acoustic features of the generalization stimulus were most similar to the features of the talker(s) heard during the learning phase. They suggested that one potential explanation for the improved generalization findings in multiple-talker conditions in the literature is that in conditions with multiple talkers, the likelihood of overlapping acoustic features is greater than in single-talker conditions. This is contrasted with the theory described above suggesting that exposure to multiple talkers helps relax category boundaries.

The results of the present study do not strongly support one notion or the other. Rather, they suggest that there may be a ‘sweet spot’ for the benefit of stimulus variability, where exposure to multiple talkers helps listeners extend the boundaries to the benefit of a talker whose speech aligns with the acoustics of the talkers they have just heard, but doesn’t extend these internal category boundaries sufficiently to accommodate a talker with an entirely new set of acoustic differences.

Inhibitory control, recognition, and rapid adaptation

In this study, individual strength of inhibitory control, as measured by the Stroop test, was found to be predictive of overall speech recognition and to influence patterns of rapid adaptation to non-native English speech. When speech recognition scores were examined in blocks representing start of adaptation, end of adaptation, and generalization, the magnitude of the Stroop effect was shown to be predictive of performance, regardless of listener group or test point. Larger magnitude Stroop effects (i.e., poorer inhibitory control) were associated with poorer speech recognition scores.

This finding is consistent with prior literature documenting a relationship between inhibitory control and speech recognition under challenging conditions (Dey & Sommers, 2015; Janse, 2012b; Sommers &

Danielson, 1999). The Stroop effect, as measured in this study, represents an individual's capacity to inhibit their automatic response to a written word in order to process and respond to the text color. When listening to non-native speech, listeners must at times inhibit an automatic misperception of a speech segment that has been produced differently than they might expect; alterations to the acoustic signal induced by non-native accent can increase the activation of lexical competitors during the speech recognition process (Porretta & Kyröläinen, 2019).

For example, the acoustic analyses performed on the stimuli from this experiment show that the non-native talkers had lower category separability between /i/ and /ɪ/ phonemes. A reduced distinction between these two phonemes could easily lead to misperception if they were the critical phoneme in a minimal pair such as SLEEPS/SLIPS. Suppose a Spanish-accented English talker produced the sentence THE BABY SLEEPS ALL NIGHT, and produced a token with an /i/ that fell closer to the formant distribution of the listener's expected /ɪ/. A listener who is more inclined to respond with their automatic perception of the acoustic information might report THE BABY SLIPS, and indicate not having heard or understood the rest of the sentence. A listener who has a greater capacity to inhibit the automatic perception of SLIPS might be better able to take advantage of the rest of the sentence, and use the lexico-semantic cues from the remainder of the sentence to support perception of SLEEPS. The integration of these top-down cues would thus facilitate an adjustment of the listener's internal category boundaries for /i/ and /ɪ/ for the target talker(s), and allow for improved recognition of future tokens. Stronger inhibition of the automatic bottom-up response would result in these listeners showing both higher overall speech recognition, and a faster rate of learning, as was seen for most listener groups and conditions.

The influence of Stroop scores on the patterns of rapid adaptation was examined for the exemplar comparison of single SPA (C2) vs multiple SPA talkers (C4). In most conditions, larger Stroop effects were again associated with poorer speech recognition scores, but the effects differed across conditions and listener groups. For YNH listeners, in the single-talker condition, Stroop effect scores primarily affected overall recognition level. In the multiple-talker condition, Stroop effect scores influenced both level and pattern, with the smaller magnitude Stroop score associated with both higher performance and maintenance of improvements later in the condition. These patterns are consistent with the predictions described above. The OHI listeners also showed differences across condition with Stroop effect contributing to differences in overall performance levels and patterns in the early and middle adaptation trials.

However, the older adults' patterns did not align as clearly. ONH listeners showed dissimilar effects of Stroop between Conditions 2 and 4, with a reversal of the expected effect in Condition 2: larger magnitude Stroop effects resulted in higher predicted performance and maintenance of improvement. The reason for this effect in Condition 2 for the ONH listeners is unclear. However, it is interesting to observe that the interactions of Stroop and trial number in Condition 2 did not emerge as significantly different between YNH and ONH listeners, despite the visually apparent reversal. It is also worth noting that these predicted performance patterns are based on the data included in the model, and that predictions at the extreme ends of the range of scores may be less reliable due to the presence of fewer data points. A future targeted investigation including larger participant groups and a broader distribution of Stroop scores is warranted.

The other individual measures of cognitive function (attention, executive function, and working memory span) were not found to be predictive of speech recognition in this study. These non-significant relationships were unexpected, as these individual factors have been documented previously in the literature (Akeroyd, 2008; Anderson, White-Schwoch, et al., 2013; Füllgrabe et al., 2015; Füllgrabe & Rosen, 2016; Ingvalson et al., 2017). However, others have documented non-significant relationships between these cognitive domains and speech recognition ability (Colby et al., 2018; Rotman et al., 2020); it is likely that differences in study methodology and in measurement tools for the cognitive domains are related to these inconsistent findings (Heinrich & Knight, 2016).

Conclusion

The goal of this study was to evaluate rapid adaptation to non-native speech, using conditions that differed in the level of variability present in the stimulus. Single-talker and multiple-talker lists of sentences produced by NNE talkers were heard by younger and older adults, with and without age-related hearing loss. The results indicate that changes to the acoustic features of the stimulus induced changes in the overall level of speech recognition performance during the course of rapid adaptation, but did not affect the time course patterns of performance. The overall-level changes were similar between the younger and older normal-hearing listeners, but were reduced for the older hearing-impaired listeners, in some cases. There were no systematic effects of stimulus variability on magnitude of adaptation, but conditions in which listeners heard multiple talkers facilitated generalization of learning to unfamiliar talkers with the same L1. Individual measures of inhibitory control predicted different patterns of rapid adaptation between a single-talker and multiple-talker condition, for all listener groups. Overall, this study suggests that increases in stimulus variability did not significantly hinder rapid adaptation to non-native English speech, and in fact benefitted listeners in generalization of learning.

Chapter 3: Study 2

Assessing the mediating effect of semantic context on adaptation to variable stimuli, and subsequent generalization to unfamiliar stimuli

Introduction

Semantic context and speech recognition

Semantic context supports speech recognition (Kalikow et al., 1977; Miller et al., 1951; Nittrouer & Boothroyd, 1990): words in isolation or in weakly constraining sentence contexts are recognized less accurately than words presented in meaningful sentences. Several studies have been conducted that compared the context benefit under conditions varying in stimulus quality. For example, Aydelott et al. (2006) found that the N400, an electrophysiologic marker of the context benefit, was delayed and reduced in magnitude when stimuli had been low-pass filtered, suggesting a reduced context benefit for degraded speech.

This dependency on the signal quality for a context benefit is also relevant when examining the effort associated with speech recognition. In one study (Winn, 2016), young adults with normal hearing were presented with high-predictability (HP) and low-predictability (LP) sentences from the R-SPIN corpus (Bilger et al., 1984). The sentences were either presented in intact form or in an 8-channel vocoded condition, causing the sentences to be spectrally degraded. The context benefit was calculated as the difference in pupil dilation patterns (which index the effort associated with sentence recognition) and repetition between the HP and LP conditions, and was defined as an “effort release.” Listeners showed effort release for both intact and noise-vocoded sentences, but in the vocoded condition, the effort release occurred later in time and with a smaller magnitude. This finding was interpreted to indicate that the presence of signal degradation reduced the benefit provided by semantic context in the HP condition.

Goy et al. (2013) examined the semantic context benefit using three different forms of signal distortion, including low-pass filtering, time-compression, and concurrent 12-talker babble. Sentence frames included one of three levels of context to cue the final target word: congruent, incongruent, or neutral. Reaction times (RTs) to a lexical decision task were explored for these stimuli in order to determine the degree of facilitation provided by contextual information. Overall, the facilitation scores were greater for the undistorted conditions than for the distorted conditions, suggesting a greater benefit of context for acoustically intact stimuli, with no significant differences across distortion types. In sum, this study showed that listeners were able to benefit from the presence of supportive semantic context on a behavioral task, though this benefit was reduced by the presence of signal distortion.

Semantic context also appears to benefit recognition of speech produced by non-native English talkers. Behrman and Akhund (2013) measured listeners’ ratings of comprehensibility and accent strength, as well as intelligibility scores, for Spanish-accented English speech produced by talkers with mild, moderate, and strong accents. They found that listeners benefitted from contextual information when listening to all three accent strength levels, but that the context effect was largest and most consistent in the strongest accent condition. Paired with the findings of Goy et al. (2013) described above, these results suggest that semantic context is beneficial across different types of signal distortions and alterations, but that the degree of signal alteration may influence the strength and direction of the context benefit.

Semantic context, stimulus type, and rapid adaptation

In addition to benefitting overall speech recognition performance, availability of lexico-semantic information is known to promote perceptual adaptation to unfamiliar speech signals. Lexical information facilitates perceptual adaptation to ambiguous phonemes in single word contexts (Eisner & McQueen, 2005; Norris et al., 2003). In this classic paradigm, listeners heard an ambiguous phoneme falling between /f/ and /s/ in the context of words that ended in either /f/ or /s/. Following this exposure, listeners who had heard the ambiguous phoneme in an /-s/ final context were more likely to categorize tokens on an /f/-/s/ continuum as /s/, than listeners who had heard the ambiguous phoneme in /-f/ final words. These findings suggest that the listeners used the lexical information present in the exposure stimuli to adjust their internal boundaries of category representation to include the ambiguous phoneme. Babel et al. (2019) examined lexically guided learning for ambiguous phoneme stimuli that had been artificially altered to fall along the s/ʃ continuum. They found that lexically guided learning was strongest for more ambiguous stimuli, and that learning was reduced for maximally or minimally altered stimuli.

Availability of lexical-semantic information has also been shown to influence perceptual learning and rapid adaptation to unfamiliar or challenging speech (Davis et al., 2005; Maye et al., 2008; Norris et al., 2003). For example, Davis and colleagues (2005) assessed the benefits of training with noise-vocoded versions of standard English sentences, semantically anomalous (but syntactically intact with real words) sentences, and Jabberwocky (syntactically intact with non-real content words) sentences. They found that listeners who trained on sentences containing lexical information (standard English, semantically anomalous, Jabberwocky) all showed greater learning than those who trained with non-words or who did not train at all. The presence of lexical information in the training stimuli appears to have been critical for learning of spectrally distorted stimuli. Similar findings have been documented for non-native speech, with both synthetic and naturally occurring accents (Cooper & Bradlow, 2016; Maye et al., 2008).

Summary and hypotheses

In this study, the relative effects and interactions of acoustic variability and semantic context on rapid adaptation to non-native speech are explored. It was expected that conditions with a greater degree of semantic information would result in a faster rate and larger magnitude of adaptation, because lexical information is known to guide adaptation (Davis et al., 2005; Norris et al., 2003; Scharenborg & Janse, 2013). Prior literature suggests that the benefit of semantic context for speech recognition is reduced when the signal is acoustically degraded (Aydelott et al., 2006; Goy et al., 2013), but that it may be strengthened for a more naturalistic form of signal alteration such as non-native speech (Behrman & Akhund, 2013; Bent et al., 2019). To further probe the effects of acoustic stimulus features on rapid adaptation, two non-native talker conditions are evaluated in addition to a native talker condition: a single NNE talker condition and a multiple NNE talker condition. This manipulation of stimulus variability is hypothesized to interact with the effect of semantic context, such that the context benefit will be maximal at the intermediate level of stimulus variability (i.e., single NNE talker) (Babel et al., 2019).

Method

Participants

Participants for this study included 365 total listeners between the ages of 18-31 years (mean 24.17 years). Participants were recruited and compensated for their time via the Prolific online recruitment platform. Listeners were required to report the United States as their country of birth and country of current residence. In addition, all participants reported learning American English as their first language, and no experience with any languages other than English before the age of 7. Further, any listener reporting regular exposure to non-native English speech from family members or caregivers during childhood was disqualified from participation. All listeners reported no hearing difficulties, no history of ear surgeries, and were required to score lower than a 6 on the Revised Hearing Handicap Inventory (Cassarly et al., 2020). Each listener was assigned to complete one of nine conditions, with approximately 40 participants per condition. Details about the participants included in each condition can be found in Table 3.1.

Table 3.1. Characteristics of the study participants. ANOM = anomalous; STD = standard; TG = topic-grouped; NE = Native English; NS = Native Spanish; ML1 = Multiple native languages (L1s).

Condition	**n (#females)**	**Age mean (*sd*)**	**Hearing handicap score mean (*sd*)**
ANOM_NE	40 (21)	25.03 (*3.64*)	0.30 *(0.85)*
ANOM_NS	40 (26)	24.28 (*3.82*)	0.15 *(0.7)*
ANOM_ML1	41 (15)	24.27 (*3.64*)	0.20 *(0.75)*
STD_NE	39 (19)	24.41(*3.61*)	0.41 *(1.04)*
STD_NS	41 (24)	24.02 (*3.66*)	0.49 *(1.25)*
STD_ML1	40 (17)	23.73 (*3.23*)	0.05 *(0.32)*
TG_NE	40 (18)	23.9 (*3.5*)	0.15 *(0.53)*
TG_NS	42 (17)	23.45 (*3.66*)	0.38 *(1.01)*
TG_ML1	42 (20)	24.43 (*3.6*)	0.33 *(0.87)*

Stimuli and procedure

Talkers. Stimuli for this experiment were produced by both NE talkers and NNE talkers with moderately strong foreign accents. Stimuli were obtained from the SpeechBox corpus database (formerly OSCAAR; Bradlow, n.d.); additional talkers were recruited from the UMD community and were recorded in the Hearing Research laboratory. Talkers were rated for accent strength (Atagi & Bent, 2013) on a scale of 1 (no accent) - 9 (very strong accent) by a group of 14 young, normal-hearing, native English listeners, with the goal of including recordings from moderately-accented talkers (i.e. ratings around 4-6/9) as stimuli. Following pilot testing, a total of 24 talkers was included in the final experiment. Of these, 14 were recorded at UMD, and 10 were obtained from the SpeechBox database (formerly OSCAAR; Bradlow, n.d.). The NNE talkers had a variety of native languages including French, Hindi, Japanese, Korean, Mandarin, Portuguese, and Spanish. All talkers were male, and the NNE talkers had a mean accent rating of 5.39/9 (SD 0.79). See Table 3.2 for details of the talkers and their characteristics.

Table 3.2. Characteristics of the talkers. ANOM = anomalous; STD = standard; TG = topic-grouped; NE = Native English; NS = Native Spanish; ML1 = Multiple L1s.

Talker ID	Database	L1	Experimental Condition	Accent rating (x/9)	Avg. Intelligibility (all sentences used)
UMD1E	UMD	ENG	ANOM_NE	1.17	99%
UMD2E	UMD	ENG	ANOM_NE	1.31	98%
UMD6S	UMD	SPA	ANOM_NS	5.06	99%
UMD5S	UMD	SPA	ANOM_NS	6.36	97%
UMD1N	UMD	HIN	ANOM_ML1	6.1	99%
UMD1M	UMD	MND	ANOM_ML1	4.74	99%
UMD3S	UMD	SPA	ANOM_ML1	5.14	96%
UMD4S	UMD	SPA	ANOM_ML1	3.56	99%
UMD3E	UMD	ENG	STD_NE	1.49	100%
UMD4E	UMD	ENG	STD_NE	1.05	99%
662	Speechbox	SPA	STD_NS	6.24	99%
837	Speechbox	SPA	STD_NS	6.48	95%
J2M	Speechbox	JPN	STD_ML1	5.48	98%
544	Speechbox	POR	STD_ML1	5.4	96%
839	Speechbox	SPA	STD_ML1	5.48	98%
UMD1S	UMD	SPA	STD_ML1	5.74	97%
E1M	Speechbox	ENG	TG_NE	1.11	100%
E5M	Speechbox	ENG	TG_NE	1.42	99%
S1M	Speechbox	SPA	TG_NS	5.89	96%
UMD9S	UMD	SPA	TG_NS	5.89	99%
K7M	Speechbox	KOR	TG_ML1	5.77	98%
F2M	Speechbox	FRA	TG_ML1	4.68	95%

UMD8S	UMD	SPA	TG_ML1	4	99%
UMD7S	UMD	SPA	TG_ML1	5	97%

Stimuli. The stimuli included BKB/HINT-type sentence sets (Bench et al., 1979; Nilsson et al., 1994) that were altered from their original form to create three levels of stimulus context, including from least to greatest amount of available semantic information: anomalous sentences, standard sentences, and topic-grouped sentences. The anomalous sentences were constructed by scrambling the keywords of the sentence corpus within grammatical type such that sentences retained their syntactic structure but were devoid of semantic information. For example, the sentence “A/The FARMER KEEPS a/the BULL” becomes “A/The DOG HELPED the POTATOES”. The standard sentence sets contained unaltered sentences presented in randomized order, and the topic-grouped sentences included unaltered sentences presented in lists organized by topic, such as ‘Food and Drink’ or ‘Transportation and Travel.’ Listeners were informed of the topic prior to the presentation of the first sentence. The sentences’ conformity to the topic categories was confirmed by pilot testing with 14 young, normal-hearing listeners. An additional round of pilot testing was conducted to confirm that all sentences and talkers used in the experiment had a similar, high level of intelligibility. Young, normal hearing listeners (5 per talker) listened to and transcribed each sentence in quiet; these intelligibility scores were used to guide the formation of the experimental lists. See Table 3.2 for the mean intelligibility of each talker’s stimuli.

Procedure. The experimental procedures were carried out using two online data collection platforms: Qualtrics and PennController (Zehr & Schwarz, 2018). First, listeners completed a headphone check screening developed by Woods et al. (2017), which confirmed that listeners were using headphones to complete the experiment, rather than listening in the sound field. The headphone check was implemented via Qualtrics. In each trial of the listening check, listeners were asked to judge which of 3 presented tones is the softest. Each tone involves stereo presentation, but one of the three tones is presented 180 degrees out of phase across channels. Thus, for listeners not using headphones, the task becomes inordinately difficult due to phase cancellation. Participants who did not pass this screening were disqualified from completing the listening experiment. Following the headphone check, listeners completed a series of questionnaires probing hearing history, language experience and accent exposure history. All listeners who passed the headphone screening and were not disqualified based on their language and accent exposure histories were then advanced to the listening experiment.

Stimuli were presented in 6-talker babble at a signal-to-noise ratio (SNR) of 0 dB. The SNR was set at this level following pilot testing with 10 young, normal hearing listeners, and was intended to avoid ceiling and floor effects. It should be noted that different listeners served in each of the pilot studies described above.

The procedures for the listening experiment were similar to those described in Experiment 1.1. In each trial, listeners heard a sentence and were asked to transcribe it to the best of their ability. Written responses were recorded and stored for scoring. Following their transcription, listeners heard the same sentence a second time spoken by the same talker, and saw the text of the sentence written on the computer screen in front of them. This explicit feedback was designed to facilitate lexically guided learning (Davis et al., 2005). In each condition, listeners heard an initial set of 30 sentences (adaptation phase), followed by an additional set of 10 sentences (generalization phase). After completing the speech tasks, all listeners completed the Stroop task, a measure of inhibitory control (Stroop, 1935).

A total of nine conditions was included in the experiment, including three levels of supportive semantic context (anomalous, standard, topic-cued), and three levels of talker type [native English (NE), native Spanish (NS), and multiple L1s (ML1)]. This set of conditions allows for an examination not just of the main effect of context, but also any potential interactions of a theorized context benefit with degree of stimulus variability. Given the lack of accent-independent generalization observed in Experiment 1, generalization to an unfamiliar accent was not tested in this experiment. However, generalization to an unfamiliar talker with a familiar accent was included. In each condition, adaptation was immediately followed by a generalization test. Listeners heard 10 sentences produced by an unfamiliar talker who shared their L1 with the talkers heard during adaptation. For the

ML1 conditions, the generalization talker's L1 was Spanish. The context level of the generalization sentences was the same as in the adaptation condition, and the structure of the trials was identical.

Statistical Analyses

Adaptation

The time course patterns of adaptation were evaluated similarly to the analyses described in Experiment 1. GAMM analyses for time course of adaptation utilizing contrast coding schemes to target the comparisons of interest within the analyses were built following the recommendations of Wieling (2018) and Soskuthy (2021). The random effects structures included random smooths of both subject and token.

Magnitude of adaptation was derived by calculating a relative change measure, comparing performance at the start and end of adaptation ($\frac{End-Start}{Start}$). For this measure, "start" and "end" consisted of the average of the first and last five trials of adaptation. The relative change measures were analyzed using multiple linear regression. Talker type and context level were evaluated, and their interaction was inspected for contribution to model fit.

Generalization

In order to examine generalization to an unfamiliar talker, a generalized linear mixed effects regression (GLMER) was constructed with proportion keywords correct included as the dependent variable (DV). A 3-level factor of Test (Start of adaptation, End of adaptation, Generalization;

reference: Generalization) was included as a predictor variable, allowing for an evaluation of the performance at generalization as an improvement relative to the start of adaptation, and as a maintenance of performance at the end of adaptation. For this analysis, “Start” and “End” included 10 trials each, in order to make a balanced comparison with the 10 trials included in the generalization phase. Talker type, context level, and their interactions were all evaluated as predictors for generalization, with random effects including participant and token.

Results

Single talker conditions

Performance patterns for the two single-talker conditions are visualized in Figure 3.1.

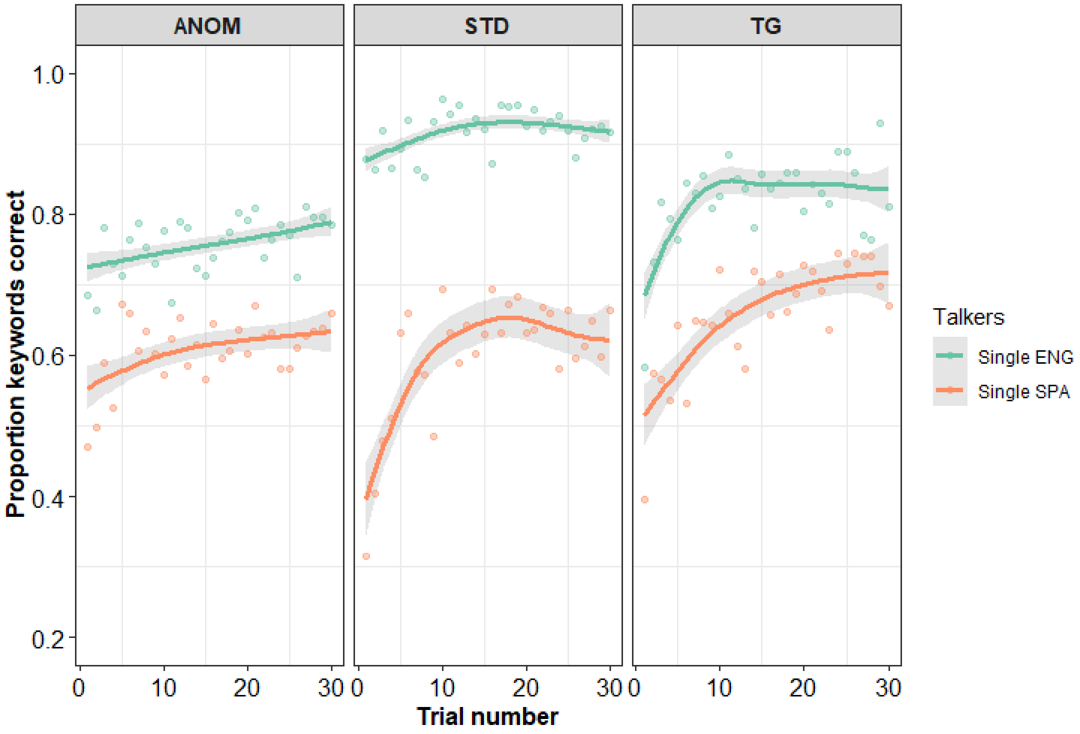


Figure 3.1. Speech recognition performance over the course of 30 trials for sentences spoken by a single native English talker (green) and single native Spanish (orange) talker, with separate panels for each level of semantic context. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. ANOM: semantically anomalous sentences; STD: standard sentences; TG: topic-grouped sentences.

The effects of talker language and context level were examined using an ordinal-coded GAMM, which allowed for consideration of both intercept and slope-related differences: i.e., were significant effects due to differences in overall performance level, or due to differences in the pattern of speech recognition across trials, or both? A full model was run including contrast-coded model terms for the effects of condition, talker, and their interactions. Random effects structure included random smooths for participant and token. The model summary is contained in Table 3.3.

Table 3.3. GAMM including ordinal terms to compare single talker conditions. Reference levels: Context = ANOM; Talker = NS. Note that an alpha of .025 is used for significance testing due to Bonferroni correction for an ordinal-coded model.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	0.55	0.18	3.11	<.01
Is_std_ordTRUE	0.01	0.25	0.03	.97
Is_tg_ordTRUE	0.40	0.25	1.60	.11
Is_ne_ordTRUE	0.79	0.25	3.16	<.01
Is_std_ne_ordTRUE	1.63	0.36	4.44	<.0001
Is_tg_ne_ordTRUE	0.21	0.35	0.59	.56

Smooth terms	***edf***	***Ref.df***	***χ2***	***p***
s(Trial)	6.14	7.18	35.18	<.0001
s(Trial): Is_std_ordTRUE	3.00	3.63	13.70	<.01
s(Trial): Is_tg_ordTRUE	3.24	3.93	14.7	<.01
s(Trial): Is_ne_ordTRUE	1.00	1.00	0.07	.80
s(Trial): Is_std_ne_ordTRUE	3.18	3.88	4.66	.25
s(Trial): Is_tg_ne_ordTRUE	1.00	1.00	1.71	.19
s(Trial, Subject)	367.15	537.00	1659.09	<.0001
s(Trial, Token)	550.77	718.00	3468.63	<.0001

Releveling was used to examine the comparisons of interest not represented in the model seen in Table 3.3. The models indicated that, at each level of context, performance was significantly lower for speech produced by the NS talker as compared to the NE talker (STD: β=2.4, SE = 0.26, z = 9.3, p<.001; TG: β=0.99, SE = 0.25, z = 3.95, p<.001; ANOM: β=0.79, SE = 0.25, z = 3.16, p<.01). Additionally, the parametric interaction of talker and context was significant for the STD vs ANOM (β=1.63, SE = 0.36, z = 4.44, p<.001) and TG vs STD comparisons (β=-1.42, SE = 0.36, z = -3.95, p<.001), but not the ANOM vs TG comparison (β=0.21, SE = 0.35, z = 0.59, p=.56). These interactions indicate that the overall talker effect was larger in the standard condition than in either of the other two conditions. This is driven by higher performance in the STD condition with the ENG talker, whereas all three SPA conditions elicited similar overall performance levels.

Performance patterns, represented by the smooth terms, did not differ significantly by talker type, within any of the context levels (STD: edf = 1.73, $\chi^2 = 2.65$, $p=.31$; TG: edf = 1.08, $\chi^2 = 2.22$, $p=.14$; ANOM: edf = 1.0, $\chi^2 = 0.06$, $p=0.8$). The patterns of adaptation did differ across context types. The pattern of adaptation to the anomalous sentences was significantly different than the patterns for both the standard or topic-grouped sentences (ANOM vs STD: edf = 3.0, $\chi^2 = 13.7$, $p<.01$; ANOM vs TG: edf = 3.24, $\chi^2 = 14.7$, $p<.01$; note these values reflect the reference levels – the SPA conditions). In Figure 3.2, the terms from the model representing the smooth condition effects for STD vs TG and STD vs ANOM at the reference level (SPA) are visualized.

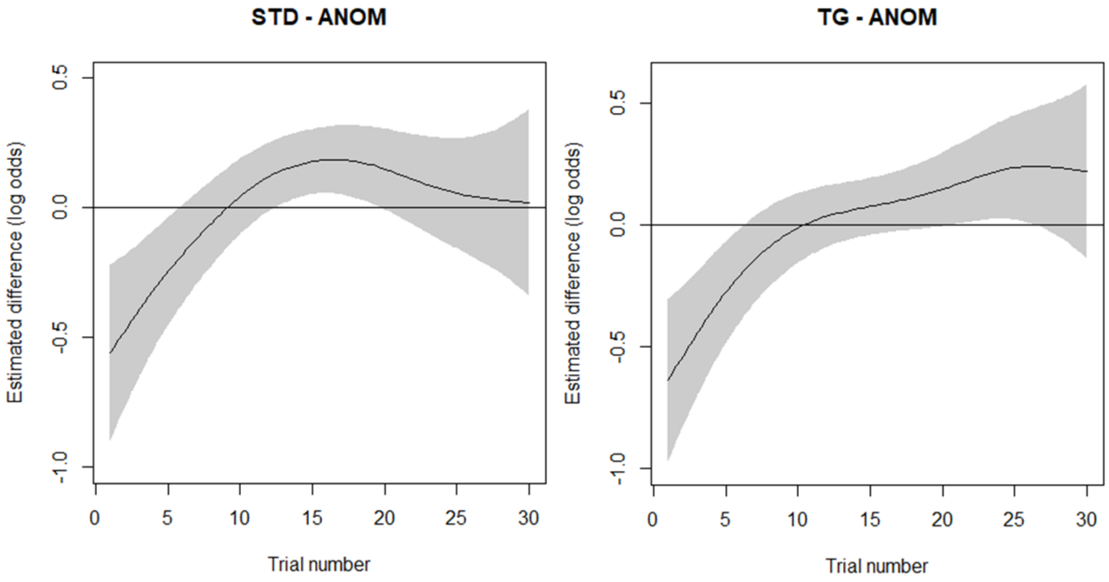


Figure 3.2. Visualization of the difference smooth terms between the standard and anomalous conditions (left) and the topic-grouped and anomalous conditions (right), for the single SPA condition. These curves represent just the differences in smooth patterns; the intercept-level differences are not visualized. Note the estimated difference is plotted in log-odds, due to the logistic modelling approach. Shading represents the 95% confidence interval; regions where the shading deviates from 0 indicate a significant difference between the patterns in the two conditions.

In these figures, the areas where the shaded regions differ from 0 represent the time ranges of significant difference. Thus, for both the standard and topic-grouped sentences, there is a significant difference in the early trials in which performance is higher for the ANOM condition, but the STD and TG conditions show rapid improvement, and continue to improve and outperform the ANOM condition in the mid and late portion of trials. It should be noted that these visualized terms represent the condition effects for the SPA talker, but the non-significant interaction smooth term indicates that these condition effects did not significantly differ for the ENG talker. The patterns of performance for the standard and topic-grouped sentences did not significantly differ from one another (STD vs TG: edf = 1.0, $\chi^2 = 1.04$, p=.31).

In summary, when listening to a single unfamiliar talker, performance was significantly lower when listening to an accented talker as compared to an unaccented talker. Speech recognition performance improved over the course of 30 sentences, but the rate of adaptation differed depending on the degree of semantic information available in the sentence. When sentences were devoid of semantic information, the rate of adaptation was more gradual than when sentences had a standard degree of semantic information. The addition of global list-wise context cues did not provide additional benefit in terms of an increased rate of adaptation.

Non-native talker conditions

Performance patterns for the two non-native talker conditions are visualized in Figure 3.3.

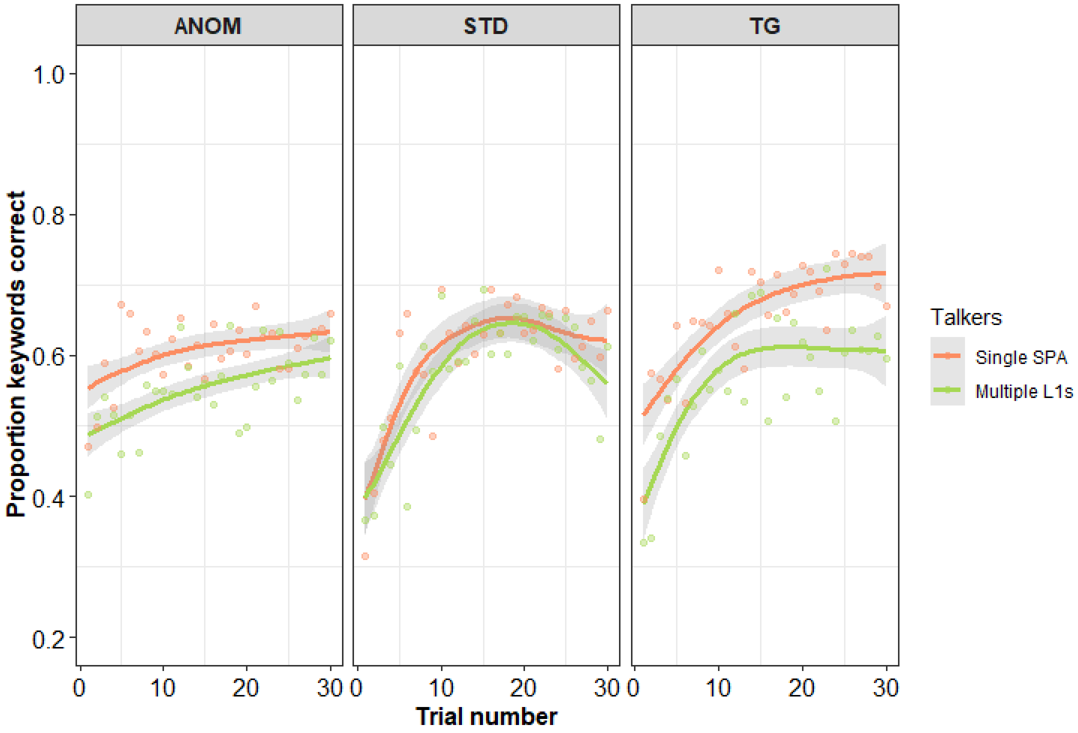


Figure 3.3. Speech recognition performance over the course of 30 trials for sentences spoken by a single native Spanish talker (orange) and by multiple talkers with unique L1s (green), with separate panels for each level of semantic context. Each point represents the raw group mean for a single trial. Lines represent the predicted values generated by the GAMM analysis, with shading representing the 95% confidence interval. ANOM: semantically anomalous sentences; STD: standard sentences; TG: topic-grouped sentences.

These two talker conditions were compared using a binary-coded GAMM to examine the effects of talker type, Context, and any potential interactions. A full model was fit including the full effects and interactions structure; random effects structure included random smooths for participant and token. Non-significant terms were removed iteratively from the model, until the final model was selected (Wieling 2021, personal communication). There was no significant difference between performance with the two talker types, nor did talker type interact significantly with context level; these terms were removed from the model. The final model included difference terms for the context effects only, and is summarized in Table 3.4.

Table 3.4. GAMM including binary terms comparing non-native talker conditions. Reference level: Context = ANOM.

	Accuracy			
Parametric coefficients	***Estimate***	***Std. Error***	***z value***	***p***
(Intercept)	0.37	0.15	2.48	<.05
Smooth terms	***edf***	***Ref.df***	***χ2***	***p***
s(Trial)	4.83	5.77	34.8	<.0001
s(Trial): Is_std_bin	4.09	4.72	16.16	<.01
s(Trial): Is_tg_bin	4.14	4.78	27.76	<.0001
s(Trial, Subject)	455.9	2175.00	2071.78	<.0001
s(Trial, Token)	542.49	2697.00	5017.38	<.0001

Examination of model terms revealed that the pattern of adaptation to anomalous sentences differed significantly from both standard sentences (edf = 4.09, χ^2 = 16.16, *p*<.01) and topic-grouped sentences (edf = 4.13, χ^2 = 27.76, *p*<.001). However, there was no significant difference in performance between the standard and the topic-grouped sentences (edf = 42.51, χ^2 = 2.43, *p*=.41) when the model was releveled to examine this comparison.

Figure 3.4 visualizes the difference smooth terms for the two significant condition effects, averaged across talker type.

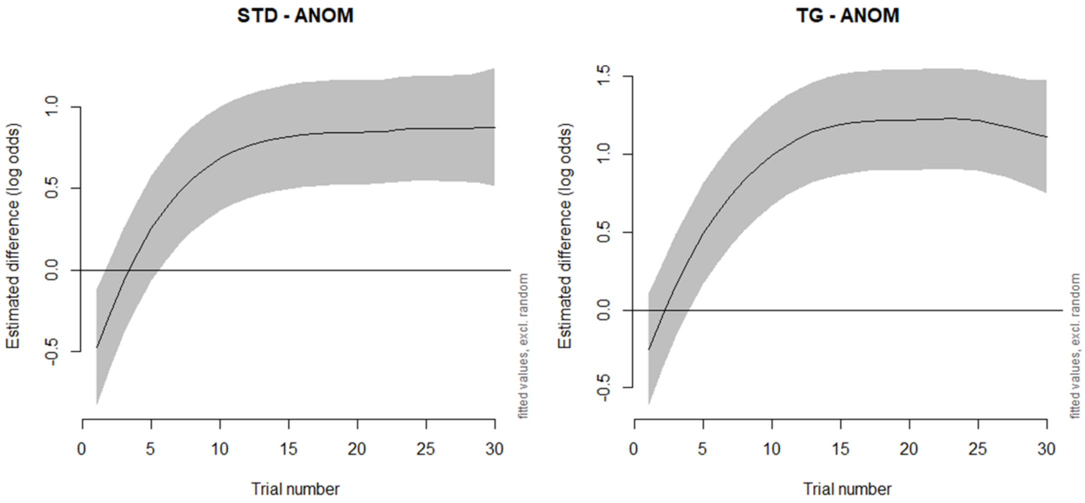


Figure 3.4. Visualization of the difference smooth terms between the standard and anomalous conditions (left) and the topic-grouped and anomalous conditions (right), averaged across talker conditions. Note the estimated difference is plotted in log-odds, due to the logistic modelling approach. Shading represents the 95% confidence interval; regions where the shading deviates from 0 indicate significant differences between performance in the two conditions.

In both cases, performance is generally similar at the outset of trials, but performance increases more rapidly in the two context-rich conditions as compared to the anomalous condition during the early trials. This difference in conditions slows during the second half of trials, with listeners in the standard and topic-grouped conditions showing a plateau in performance. These difference curves represent the differences in performance regardless of talker type, as talker type had been dropped from the model as a non-significant predictor.

Individual listeners' scores on the Stroop task were evaluated for contribution to the variance on rapid adaptation performance; inclusion of Stroop scores did not significantly contribute to model fit (*p*=0.99).

Magnitude of adaptation

Magnitude of adaptation was calculated by comparing performance on the initial and final 5 trials of adaptation. This relative change measure is plotted in Figure 3.5, and was compared across conditions.

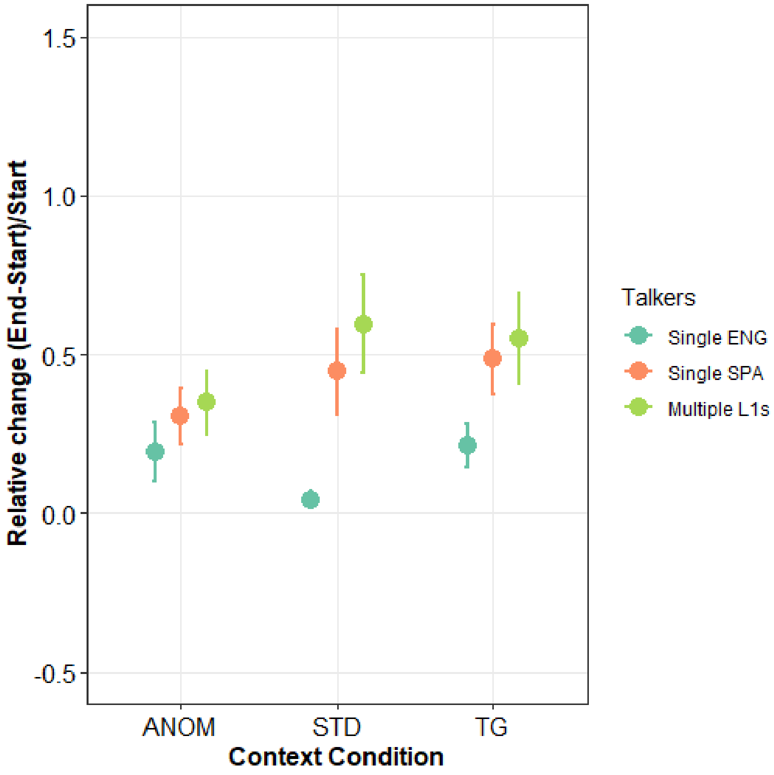


Figure 3.5. Magnitude of rapid adaptation observed in each condition. Outlier values falling outside of 2 standard deviations from the mean were removed from plots and analysis. Error bars reflect standard error. ANOM: semantically anomalous sentences; STD: standard sentences; TG: topic-grouped sentences.

Seven outlier values were removed from the analysis. A linear regression was fit to the data examining the effects of talker type, context level, and their interactions. The final model selected (F(2, 355) = 8.31, R^2 = 0.04, $p<.001$) contained only a main effect of talker type. The main effect of context, and the interaction of talker type and context, did not contribute significantly to model fit (*p*>.05, all comparisons). The individual Stroop effect scores were examined for contribution to model fit, but inclusion of Stroop did not improve the model (*p*>.05). The final model output can be found in Table 3.5.

Table 3.5. General linear regression for magnitude of adaptation. Reference level: Talker = NE.

	Relative Change			
Predictors	*Estimate*	*std. Error*	*t-value*	*p*
(Intercept)	0.15	0.06	2.40	<.05
Stim [NS]	0.26	0.09	2.96	<.01
Stim [M.L1]	0.35	0.09	3.91	<.001
Observations	358			
R^2 / R^2 adjusted	0.045 / 0.039			

The effect of talker type indicates that the magnitude of adaptation was greater for both NNE talker conditions as compared to the NE condition (NE vs SPA: β = 0.26, SE = 0.09, *t* = 2.96, *p*<.01; NE vs ML1: β = 0.36, SE = 0.09, *t*= 3.91, *p*<.001). The magnitude of adaptation did not differ significantly between the two NNE conditions (SPA vs ML1: β = 0.08, SE = 0.09, *t* = 0.95, *p*=.35). However, it was also noted that the magnitude of adaptation for the NE talker was significantly greater than 0 (β = 0.15, SE = 0.06, *t* = 2.4, *p*<.05), meaning that improvements in performance over trials was seen even for the NE talker. In sum, magnitude of adaptation was greater for the non-native talker conditions than for the NE condition, but this was not influenced by context level.

Generalization to an unfamiliar talker

Performance on the generalization task was compared to performance at both the starting and ending points of the adaptation period. Speech recognition scores for the first and last 10 trials of adaptation were averaged, and are plotted in comparison to the 10 generalization trials in Figure 3.6.

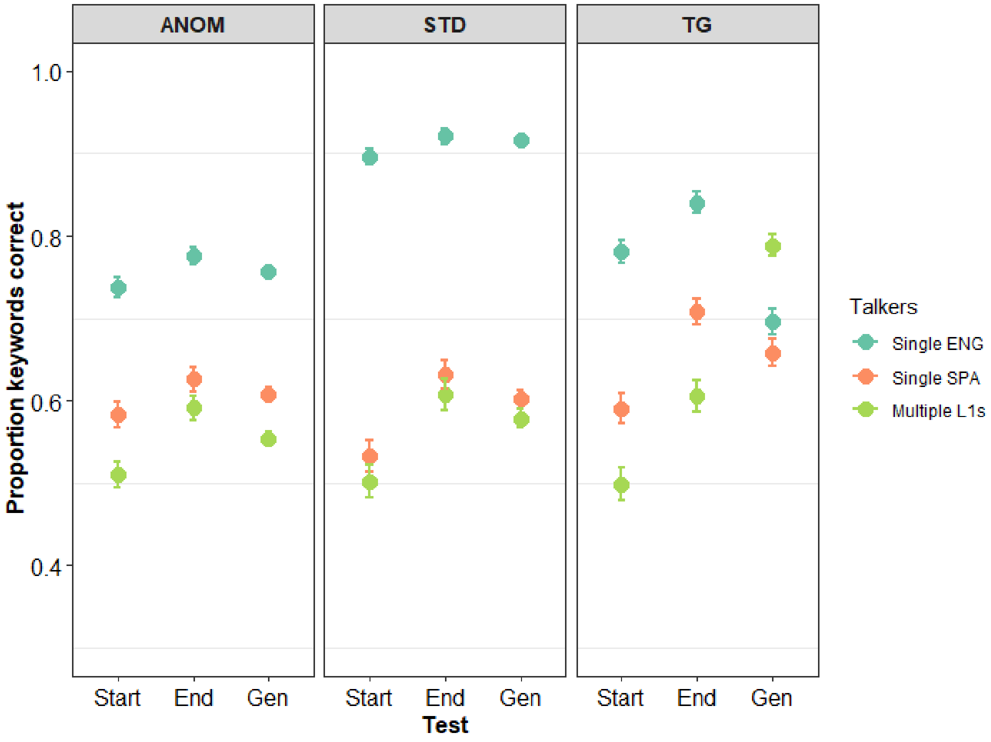


Figure 3.6. Generalization to an unfamiliar talker speaking with a familiar accent. Data displayed include the first 10 trials of adaptation (Start), the last 10 trials of adaptation (End), and the generalization phase (GF). Error bars indicate standard error of the mean. ANOM: semantically anomalous sentences; STD: standard sentences; TG: topic-grouped sentences.

The speech recognition scores for these three time points were fitted to a GLMER using a weighted binomial distribution and forward-selection procedures following the recommendations of Hox et al. (2010) in order to examine the effects of talker type, context level, and various individual predictors. The final model selected to describe the data was:

$Accuracy \sim Test * Talker \ * Context + (1|Subject) + (1|Token)$. The model summary is presented in Table 3.6.

Table 3.6. Generalization to an unfamiliar talker. NS = Native Spanish; ML1 = Multiple L1s; STD = Standard; TG = Topic-grouped; ICC = Intraclass correlation coefficient

	Accuracy			
Predictors	***Odds Ratios***	***std. Error***	***z-value***	***p***
(Intercept)	3.92	0.76	7.06	<.001
Test [Start]	0.85	0.06	-2.37	<.05
Test [End]	1.19	0.08	2.49	<.05
Stim [NS]	0.46	0.12	-2.87	<.01
Stim [M.L1]	0.28	0.08	-4.27	<.001
Context [STD]	6.00	1.67	6.43	<.001
Context [TG]	0.87	.31	-.40	.69
Test [Start] * Stim [NS]	1.00	.10	.02	.99
Test [End] * Stim [NS]	0.94	.09	-.62	.54
Test [Start] * Stim [M.L1]	0.98	.09	-.20	.84
Test [End] * Stim [M.L1]	0.93	.09	-.75	.45
Test [Start] * Context [STD]	0.89	.11	-.95	.34
Test [End] * Context [STD]	0.96	.13	-.33	.74
Test [Start] * Context [TG]	1.87	.64	1.81	.07
Test [End] * Context [TG]	2.16	.75	2.22	<.05
Stim [NS] * Context [STD]	0.17	.07	-4.56	<.001
Stim [M.L1] * Context [STD]	0.32	.13	-2.73	<.01
Stim [NS] * Context [TG]	1.56	.80	.87	.38
Stim [M.L1] * Context [TG]	5.18	2.70	3.15	<.01
(Test [Start] * Stim [NS]) * Context [STD]	0.93	.15	-.45	.65
(Test [End] * Stim [NS]) * Context [STD]	1.07	.18	.41	.68
(Test [Start] * Stim [M.L1]) * Context [STD]	0.92	.14	-.51	.61
(Test [End] * Stim [M.L1]) * Context [STD]	1.26	.21	1.43	.15
(Test [Start] * Stim [NS]) * Context [TG]	0.48	.23	-1.50	.13
(Test [End] * Stim [NS]) * Context [TG]	0.64	.31	-.90	.37
(Test [Start] * Stim [M.L1]) * Context [TG]	0.19	.10	-3.28	<.01
(Test [End] * Stim [M.L1]) * Context [TG]	0.26	.13	-2.73	<.01
Random Effects				
σ^2	3.29			
$\tau_{00\ token}$	1.54			

$\tau_{00\ \mathrm{Name}}$	0.41
ICC	0.37
N_{Name}	361
N_{token}	550
Observations	15748
Marginal R^2 / Conditional R^2	0.135 / 0.457

The significant three-way interaction of test point, talker type, and context level was examined using the emmeans package. The *post-hoc* comparisons showed that, in the anomalous and standard conditions, performance at generalization was consistently higher than at the start of adaptation ($p<.05$, all comparisons). In the topic-grouped conditions, performance at generalization was only higher than the start of adaptation for the multiple talker condition ($\beta = -1.2$, SE = 0.35, *z-ratio*: = 3.37, $p<.01$); for the single talker conditions the difference was not significant ($p>.05$, both comparisons). When comparing generalization with the end of adaptation, performance was typically stable, with the exception of three conditions where performance at generalization was significantly lower (ANOM_NE, STD_ML1, TG_NE, $p<.05$, all comparisons).

This *post-hoc* analysis also allowed for an alternate measure of adaptation; performance was significantly higher at the end of adaptation than at the start for all conditions tested ($p<.05$, all conditions). This finding supplements those seen in the nonlinear time course analyses above: recognition improved over the course of 30 trials for all combinations of context level and talker type.

In summary, listeners' performance improved during the course of adaptation and, in most conditions, listeners were able to maintain these improvements when tested on a new talker. The two conditions where listeners did not maintain improved performance at generalization were the two single-talker conditions with topic-grouped stimuli. One possible explanation for this finding is that, in the single-talker conditions, the supra-sentence context allowed listeners to rely entirely on the top-down information for recognition, rather than using the contextual information to help adjust their internal category boundaries for processing the acoustic input. The acoustic challenge posed by the multiple-talker condition may have been sufficient to trigger some adjustment to the bottom-up processes for recognition that carried over to the generalization phase in this condition.

Discussion

The goal of this study was to evaluate two stimulus-related factors for their effects on rapid adaptation to unfamiliar speech. A series of listening conditions were evaluated that varied the level of semantic context available to the listener, as well as the type and number of talkers. It was expected that increasing the level of semantic context would facilitate more rapid adaptation, and that increasing stimulus variability would reduce or slow the adaptation process. It was also hypothesized that semantic context and stimulus variability would interact such that the benefit of context would be greatest for the intermediate level of stimulus variability, i.e., when listeners heard a single native Spanish talker.

Effects of talker type on recognition and adaptation

When the two single-talker conditions were compared, a clear difference in overall speech recognition performance was observed. Listeners had better speech recognition ability for a single native English talker than for a single native Spanish talker. This finding is consistent with prior literature; the presence of a non-native accent is known to inhibit speech recognition, especially in the presence of competing talkers (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2010a, 2010b). Non-native speech contains alterations to the acoustic features of the speech stimuli that can lead to misperceptions (Flege, 1988). Fortunately, young adults are able to quickly adjust to an unfamiliar non-native accented talker, as is seen in this and prior studies (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009b).

Performance on the single native Spanish condition was also compared with a multiple-talker condition, in which all talkers had different language backgrounds. In many prior reports, speech recognition and recall are lower for conditions containing multiple vs single talkers (Goldinger et al., 1991; Mullennix et al., 1989; Nygaard et al., 1995; Sommers et al., 1994); this effect was not seen in the present study. There were no significant differences in overall level of performance between the NS and the ML1 conditions. One possible explanation for the lack of an effect of stimulus variability may be that both conditions contained stimuli from non-native, accented talkers. The classic reports cited above of the detrimental effect of talker variability on speech recognition compare performance with native talkers. Perhaps the challenge imposed by NNE speech is more salient than that imposed by multiple talkers, when all three talkers have similar intelligibility and ratings of accent strength.

A few prior studies have specifically examined the effects of single vs multiple talkers when measuring rapid adaptation to non-native speech. Bent and Holt (2013) compared word identification performance with single and multiple (n=4) NNE talkers and found a detriment of multiple talkers. However, their study used individual word stimuli whereas the present study utilized sentence-length stimuli, which are known to elicit generally higher performance. They also varied gender (held constant in the current study) which is an additional source of stimulus variability. Kaplowicz et al. (2018) included conditions comparing performance with a single NNE talker vs five NNE talkers and found that performance was lower for the multiple talker condition. This prior study utilized IEEE sentences (IEEE, 1969), and mixed genders in the multiple-talker conditions. It may be that the stimuli used in the present study (HINT/BKB sentences) were relatively less challenging than those presented in the two prior studies, and/or that the degree of variability was greater in the prior studies, which had more talkers and mixed genders, increasing variability both in terms of acoustical and indexical features.

Effects of semantic context on recognition and adaptation

Across all three levels of context, listeners showed improvements in speech recognition performance over the course of the 30 sentences. However, the patterns of improvement differed depending on context level.

When sentences were syntactically correct but devoid of any semantic meaning, the pattern of adaptation was linear, and relatively shallow. This pattern contrasted with those seen for standard and topic-grouped sentences. In these conditions, listeners showed a steep initial increase in performance, with a plateau and/or shallower improvement in the second half of trials. These performance differences were seen across all three talker types.

This general finding aligns with the literature indicating that lexical and contextual information support perceptual learning for challenging speech stimuli (Davis et al., 2005; Hervais-Adelman et al., 2011; Jesse & McQueen, 2011; Maye et al., 2008; Norris et al., 2003). In these prior studies of lexically guided learning, there is evidence that the degree of semantic context available does not influence rate or magnitude of learning, provided that some meaningful lexical information in the listener’s native language is available (Cooper & Bradlow, 2016; Davis et al., 2005; Luthra et al., 2020). While all conditions in this study did contain intact lexical information [i.e. this study did not contain non-word conditions, as seen in studies by Davis et al (2005) and Cooper and Bradlow (2016)], rate of learning was shown to be slowed in conditions where the sentences were semantically anomalous.

It was hypothesized that this study would find not only a benefit for semantically rich sentences as compared to semantically anomalous sentences, but an additional advantage of supra-sentence context in the form of topic-grouping; this was not seen for any of the talker conditions. One possible explanation for this lack of additional benefit may be related to the sentences used as stimuli in this experiment. The HINT/BKB sentences are relatively simple, declarative sentences that contain a relatively high level of internal contextual information as compared to more challenging corpora such as the Harvard IEEE sentences (IEEE, 1969). Thus, the topic groupings may not have been beneficial in boosting learning for these simple sentences.

The additional acoustic challenges imposed by the non-native talker(s) utilized in this experiment were expected to increase listeners' reliance on contextual information in this study, resulting in an interaction of context and talker type. This interaction was not observed, contrasting with prior findings of an interaction of context level and stimulus quality (Aydelott et al., 2006; Goy et al., 2013; Winn, 2016). However, the alterations to stimulus type in the present study differ from those in prior studies; here, listeners were tested on non-native English speech, while in prior studies the listeners heard low-pass filtered, vocoded, and time-compressed speech. These results suggest that the context benefit is similar when listening to non-native speech, a naturalistic form of signal alternation, as compared to native English speech. The study additionally contrasts with the findings of Behrman and Akhund (2013) and Bent et al. (2019), who showed an increased context benefit for NNE speech, particularly in stronger accent conditions. In the present study, the NNE talkers were selected to have a similar, moderate level of accent strength. Thus, the contrast between single and multiple NNE talkers may not have been significantly detrimental to trigger an increased reliance on semantic context for recognition and learning.

Magnitude of adaptation

In addition to examining the time-course patterns of rapid adaptation, this study measured the magnitude of adaptation to each context and stimulus variability manipulation. Magnitude was measured as a comparison of performance between the first and last 5 trials of the adaptation condition. The analyses showed that talker type had an effect on magnitude of adaptation, with the native talker condition eliciting a smaller magnitude than either of the non-native talker conditions. This was true regardless of the level of context. This finding likely relates to the overall higher level of speech recognition performance for the native as compared to the non-native talker conditions. In conditions where starting performance is lower, listeners show greater magnitude of learning and adaptation. This effect of starting level on auditory learning and adaptation has been observed by others for rapid speech (Manheim et al., 2018) and foreign-accented speech (Banks et al., 2015; Tzeng et al., 2016).

Generalization

Generalization to unfamiliar talkers was examined in this study. While the benefit of lexico-semantic information for perceptual learning of speech is well-documented (Cooper & Bradlow, 2016; Davis et al., 2005), it was unclear whether the level of context present in the adaptation stimulus would influence the degree of transfer of learning. The literature indicates that exposure to multiple talkers during an adaptation phase is beneficial in facilitating generalization, above adapting to a single talker (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Sidaras et al., 2009).

In each condition, generalization was tested with an unfamiliar talker who shared a language background with a talker heard during adaptation; the level of context was constant between adaptation and generalization. The findings of the generalization analysis indicate that generalization of learning was dependent on both the level of context and talker type. In the anomalous and standard sentence conditions, listeners performed significantly better at generalization than at the start of adaptation, regardless of talker type. In the topic-grouped sentences, this was only seen for the multiple talker condition; for the single talker conditions, there were no significant differences in performance between start of adaptation and generalization. Start of adaptation and generalization both constitute the first exposure to an unfamiliar talker; higher performance at generalization indicates that learning occurred and was maintained in these conditions, at least to some degree. Generalization to an unfamiliar talker in the anomalous conditions suggests that although the rate of learning was slower in these conditions, the learning experience was still sufficient to facilitate relatively high recognition of an unfamiliar talker with a familiar accent. Thus, the retuning of internal category boundaries for mapping acoustic input to lexical meaning that occurred during adaptation was slowed by the lack of contextual information, but was not eliminated altogether.

Conclusion

This study evaluated the relative contributions of semantic context and stimulus variability on rapid adaptation and generalization to non-native English speech in young adults with normal hearing. Listeners showed a slowed rate of adaptation to semantically anomalous sentences, but the magnitude of learning and generalization to unfamiliar non-native English talkers in this context condition was not reduced as compared to the semantically intact conditions. Overall speech recognition performance was lower in non-native talker conditions than in native talker conditions, but patterns of adaptation were similar between native and non-native talkers. Together, these results indicate that manipulations of bottom-up acoustic detail influenced overall performance levels, and contextual manipulations affected the time-course of adaptation, but that these two effects were independent of one another.

Chapter 4: Study 3

Examining the effects of semantic context and stimulus variability on electrophysiologic measures of lexical access

Introduction

The relative success or failure of a spoken communication encounter can be highly influenced by factors related to the listener, talker, or stimulus. In laboratory settings, these factors can be experimentally manipulated while speech recognition is assessed, often via behavioral tasks such as repetition or transcription. While these behavioral tasks are valuable in illuminating listener perception, repetition-based measures are more limited in their ability to answer questions about the processes underlying speech recognition and perception. Objective methods such as eye-gaze measures and electrophysiology have therefore been critical in expanding the understanding of speech recognition processes from cochlea to cortex.

Electrophysiologic measures of a semantic context benefit

The N400 component is an event-related potential (ERP) measure commonly used to observe the effects of semantic context on speech recognition. The N400 component is a centrally distributed, negative-going potential occurring around ~300-500 ms, and is thought to index the relative ease of lexical access and semantic integration (Lau et al., 2008; 2009). The negative-going nature of the N400 component can cause confusion when discussing the relative size of the effect. A *larger* N400 component will have a significantly more *negative* absolute amplitude, typically a negative voltage value. The magnitude of the N400 deflection is directly connected to ease of lexical access (Federmeier & Kutas, 1999; Lau et al., 2009). Specifically, the N400 component in response to a target item that is relatively more difficult to map to a stored lexical representation will have a larger (i.e. more negative) amplitude than one in response to an item that is easier to access. Experimental conditions can be compared by examining the N400 *effect*, which is derived by calculating a difference potential: typically, a subtraction of the lexically “easier” condition from the lexically “harder” condition. This can be visualized as a difference wave, where the effect is seen as a negative-going deflection centered around 400 ms after onset of the target stimulus.

Many factors can contribute to the relative ease or difficulty of lexical access. The standard example is a comparison of conditions in which target words are presented within sentence contexts with varying levels of cloze predictability (a measure indicating the probability that the sentence frame will be completed with that word), i.e. an *N400 effect of context*. An example of this contrast is “I like my coffee with cream and SUGAR/DOGS” (Federmeier et al., 2003; Kutas & Hillyard, 1980). A word with lower cloze probability (here, DOGS) would require relatively greater resources to access, thus resulting in an N400 component with a greater magnitude of deflection. The magnitude of the N400 component is thought to correspond with degree of semantic expectancy. The semantic or contextual information can serve to narrow a listener’s expectations about upcoming lexical items. If the target item violates these expectations, the result is a need for increased processing resources, which result in the greater N400 magnitude. This has been illustrated by a number of studies demonstrating that the magnitude of the N400 component is modulated by the degree of target word predictability within a sentence or in relation to a semantic prime: within-category semantic violations have less of an effect on the N400 amplitude than across-category violations (Federmeier et al., 2010; Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). For example, Federmeier and Kutas (1999) compared the EEG responses of young adults to the final word of the sentence: “They wanted to make the hotel look more like a tropical resort, so along the driveway they planted rows of PALMS/PINES/TULIPS”. In this case, PINES represents a within-category violation, while TULIPS represents an across-category violation. They found that the N400 deflection was greatest for TULIPS and minimal for PALMS, with PINES falling in between. Latencies of the N400 component are also outcomes of interest; a late N400 component suggests a delayed or inefficient integration process.

Lexical access and signal type

Another potential contributor to the difficulty of lexical access is clarity of the target signal. Degradations or alterations to signal quality are thought to impact the process of lexical activation, and thus influence the N400 amplitude. These changes to the signal can result from naturally occurring sources (i.e. non-native accent), or artificial manipulations (i.e. time-compression, noise-vocoding). Influences of signal alteration have also been demonstrated for ERP components reflecting earlier stages of the speech recognition process, including the N1 and P2 components, which correspond to auditory object detection and feature extraction (Straus et al., 2013).

A small number of studies have examined the effect of non-native talker status on N400 amplitude in the absence of explicit context or predictability manipulations: i.e., an *N400 effect of talker*. Goslin et al. (2012) had participants listen to low predictability sentences that were produced by either a native talker or talkers with a regional or foreign accent. They found that the N400 component was largest in magnitude for the foreign-accented speech, with no differences in deflection magnitude between the native and regional-accented speech. The authors concluded that the acoustic alterations imposed by the non-native accent had not been fully normalized by the listeners in early stages of processing, and thus still had an influence at the point of lexical access. In a study of recognition and adaptation to non-native speech, Romero-Rivas et al. (2015) presented participants with sentences produced by a variety of non-native talkers. They found that responses to the non-native speech showed an N400-like component with significantly greater magnitude than the responses to the native speech, indicating that lexical access was more challenging for the non-native speech. Interestingly, this N400 effect of talker decreased in magnitude over the course of the experiment. The authors interpreted this finding to reflect rapid adaptation to the non-native speech signal and increasing ease of lexical access with additional exposure to the non-native speech.

Lexical access, signal type, and semantic context

A number of studies have examined the interactions of signal alteration and semantic context on the N400 effect, i.e. the interactions of N400 effects of context and effects of talker. For example, Aydelott et al. (2006) compared ERP responses in young adult listeners to target words that were either semantically congruent or incongruent to a carrier sentence. The carrier phrases were presented as clean speech or in a low-pass filtered form. Listeners showed robust N400 effects of congruency in the clean speech conditions, but the effect was not present for stimuli that had been low-pass filtered. A later study by Straus et al. (2013) examined N400 responses to final words in sentences with variation both in terms of level of sentence context and word typicality (i.e. category violation or alignment – see the PALMS/TULIPS/PINES example on p. 134). The study included 2 levels of spectral degradation to the signal: 4-channel and 8-channel noise-vocoding. N400 effects were calculated for both the typicality effect, and a combined effect of typicality and context. In the most strongly degraded condition (4-channel vocoding), no significant N400 effects were evident. A comparison of the responses to unprocessed and to 8-channel vocoded speech showed that the N400 effect strength differed by condition when speech was clear, but was similar across the context and typicality manipulations when the signal was spectrally degraded. It was also noted that the N400 effect was delayed in latency when the signal was degraded. These findings suggest that predictive processing becomes constrained under the limitations of a distorted signal. That is, while listeners can use contextual information to generate predictions in both degraded and non-degraded listening conditions, the disadvantage afforded by a low-predictability target is reduced if the sentence frame has been acoustically degraded.

A few prior studies (Gosselin et al., 2021; Grey & van Hell, 2017; Hanulíková et al., 2012; Romero-Rivas et al., 2015) have examined the effects of both semantic information and non-native speech on the N400 response, and have found inconsistent results. Hanulíková et al. (2012) examined EEG responses to sentences with and without semantic violations in a group of young Dutch listeners. The sentences were produced by native Dutch speakers and native Turkish speakers. In this study, N400-like effects were present for both native and non-native speech when comparing sentence types, though the distribution of the effect across electrodes was broader for non-native speech as compared to native speech.

Romero-Rivas et al. (2015) presented non-native speech to young, normal-hearing listeners. The researchers found that the N400 effect elicited by semantic violations was greater in magnitude and had a broader distribution across channels for the non-native speech, compared to native speech. This finding of a more broadly distributed response, consistent with that found by Hanulíková et al. (2012), could be interpreted to reflect a recruitment of additional cognitive resources for processing non-native speech.

A subsequent study by Grey and Van Hell (2017) reported different findings. In this study, young, native English-speaking listeners heard English sentences produced by native English and native Mandarin speakers. The sentences were either well-formed or contained semantic violations. ERP responses to the spoken sentences containing semantic violations showed typical N400 responses for the native English speakers, but the responses to the non-native speech did not show an N400-like response. Rather, there was a late frontal negativity present between 500-900 ms for the non-native speech containing semantic violations. Interestingly, a *post-hoc* analysis of individual differences showed that listeners who were able to correctly identify the non-native accent were more likely to show an N400-like response pattern, while those who could not identify the accent showed the late negativity. The authors suggest that the late negativity may represent a substantially delayed lexical access process, or that listeners employ different strategies in processing semantic errors in non-native speech than in native speech.

In a recent study, Gosselin et al. (2021) probed whether the influence of talker accent on a semantically elicited N400 was dependent on error type. Listeners heard two types of violations: one that was commonly produced by non-native talkers, and one that was not. They found no differences between the N400 effect in response to native and non-native speech when examining the traditional time window of 350-600 ms, but did note that the effect persisted longer in response to the non-native speech, regardless of error type.

Overall, these studies paint an unclear picture of the effect of non-native speech on a semantically elicited N400 effect. In studies of the combinatory effects of signal alteration and semantic content on the N400 effect, purely spectral changes to the signal (low-pass filtering, noise-vocoding) seem to diminish the N400 effect, but a more global, temporal-spectral change to the signal (non-native accent) does not affect the N400 response in a predictable manner. In the studies that use non-native speech, findings include reduction (Grey & van Hell, 2017), magnification (Romero-Rivas et al., 2015), and no change to the magnitude of the response (Hanulíková & Weber, 2012), with two studies reporting a broader distribution of the response (Hanulíková et al., 2012; Romero-Rivas et al., 2015). One potential explanation for the discrepancies within the non-native accent studies may be the interplay between listener characteristics and talker accent. Grey and Van Hell (2017) intentionally recruited listeners who had minimal experience with languages other than English, whereas Hanulíková et al. (2012) reported that the majority of their participants were able to correctly identify the non-native accent. Indeed, Grey and Van Hell (2017) noted different EEG response patterns between participants who could and could not identify the accent, though Gosselin et al. (2021) did not find an effect of accent familiarity on N400 component amplitude. The indexical information, which is an inherent feature of non-native speech, may also contribute to the differences in findings between studies using non-native speech and those using more controlled, artificial forms of degradation (Aydelott et al., 2006; Straus et al., 2013).

Auditory object formation and non-native speech

Romero-Rivas and colleagues (2015) also explored the effects of non-native accent on an earlier neural response, the P200. This response, also known as the P2, is a positive-going deflection occurring around 200 ms that originates from activation of the primary auditory cortex. The P200 is understood to reflect the early stages of auditory processing, such as auditory feature detection and object formation (Reinke et al., 2003; Tremblay et al., 2001). For example, the N1-P2 complex has been shown to be sensitive to temporal speech cues such as voice-onset timing (Dimitrijevic et al., 2013; Steinschneider et al., 1994). Enhancements in the amplitude of the P200 have also been observed following auditory training (Atienza et al., 2002, Tremblay et al., 2001, 2014). In the Romero-Rivas et al. (2015) study, young adult listeners showed reduced P200 amplitudes in response to non-native as compared to native speech. This finding was interpreted to indicate a greater difficulty in processing the acoustic features of non-native speech, occurring even before higher-level lexical processing. This talker effect on P200 amplitude remained constant over the course of the experiment, which was completed in a single session.

Lexical access, semantic context, and aging

The literature cited above only investigated responses in young listeners with normal hearing. The ERP literature documenting a detriment of aging on the ability to benefit from semantic context includes studies of both auditory and visual language processing. In an early study utilizing auditory stimuli, Federmeier et al. (2002) presented older and younger listeners with target words in the final position of high-constraint and low-constraint sentences. When the N400 responses to the expected target word were examined, younger listeners showed a facilitative effect of sentence context; expected words within a highly constraining context showed a smaller N400 component amplitude than expected words within a low-constraint sentence. Older listeners did not show this same pattern: In older listeners, the N400 component in response to expected target words was similar regardless of sentence constraint. This finding suggests that the older adults were not able to use the highly constraining sentence contexts to generate predictions about the upcoming words and facilitate processing. Supporting these findings, Federmeier et al. (2003) showed that higher-level semantic constraints imposed by a spoken sentence took longer to process in older adults than younger adults.

Evidence of weaker predictive processing in older adults is also evident in responses to visually presented language. Older adults show a delay in the peak latency of the N400 effect for both sentence-final target words and semantic prime pairs (Federmeier et al., 2010; Federmeier &

Kutas, 2005). N400 amplitudes are similar between older adults and young adults for words that have low probability or are semantically incongruous, but older adults fail to show facilitation in the corresponding high cloze or high typicality conditions (Federmeier et al., 2010; Federmeier & Kutas, 2005).

Collectively, these findings suggest that older adults are less efficient and effective at making use of semantic context to generate predictions about incoming stimuli. This age-related reduction in efficient use of context is in line with literature indicating that older adults are less able to inhibit activated lexical items in order to recognize a target accurately and efficiently (Hartman & Hasher, 1991; Sommers & Danielson, 1999; Taler et al., 2010). However, much behavioral literature suggests that older adults benefit equally or more from contextual information than younger adults (Dubno et al., 2000; Goy et al., 2013; Pichora-Fuller et al., 1995; Sheldon et al., 2008; Sommers & Danielson, 1999). Though the seeming increase in contextual benefit with aging described in the behavioral literature may not reflect an increasing strength in predictive processing as seen in the ERP research, the context benefit has not yet been examined using both behavioral and electrophysiologic measures in the same individuals. A study using eye-tracking to simultaneously evaluate the context benefit via online and offline speech recognition processes in younger and older adults found an age-related detriment in processing of sentence-level semantic cues, despite an overall similar level of behavioral performance across age groups (Harel-Arbeli et al., 2021). In the present study, both behavioral and EEG measures are combined in order to elucidate the level of processing at which age effects manifest in speech recognition.

Another methodological strategy that may help shed light on some of the conflicting prior findings regarding the context benefit in older versus younger adults is to not only examine the average performance across conditions, but to examine the time-course of performance within a condition or experiment. An example of this strategy can be seen in the study completed by Romero-Rivas et al. (2015), which revealed that, in younger adults, the N400 effect of talker was reduced in magnitude with additional exposure to the talker. Examination of this time-course data can provide information about rapid adaptation, which is thought to reflect the early processes of perceptual learning. As seen in the study by Romero-Rivas et al. (2015), it is possible that some effects or interactions are present during only a portion of trials and shift as listeners adapt to the stimuli; looking at average data for entire conditions may mask some findings. In the present study, time-course data for both behavioral and electrophysiologic findings are analyzed.

Individual characteristics

Listener-related factors independent of age may also influence the effects of context and talker language background on speech recognition. Another individual factor that may influence speech recognition is the individual's cognitive capacity. The relationship between cognitive abilities and speech recognition ability has been explored extensively, though there are still significant gaps in knowledge. The Ease of Language Understanding

(ELU) model (Rönnberg et al., 2008), a model of speech recognition, posits that cognitive functions play an important role in facilitating speech understanding in challenging environments, with an emphasis on the importance of executive functions, especially working memory. The ELU model has been updated in recent years to include consideration of other aspects of executive function that are critical for speech recognition, including inhibition (Rönnberg et al., 2013).

Working memory represents the capacity to store and manipulate information, and often emerges as a significant predictor of individual performance for speech recognition, including speech in noise (Akeroyd, 2008; Anderson, White-Schwoch, et al., 2013; Füllgrabe et al., 2015) as well as non-native speech (Banks et al., 2015; Lev-Ari, 2014). The utility of the processes associated with working memory (i.e. processing and storage) in the context of speech recognition is clear: a listener must retain and manipulate acoustic-phonetic and linguistic information in order to successfully participate in spoken conversation, which necessarily involves a rapid rate of incoming sensory information. Inhibition is the process by which the undesired allocation of processing resources to non-relevant cues is prevented. The hypothesized role of inhibitory mechanisms in speech recognition is in preventing information that is irrelevant to the target from taking up resources that would be used for processing the target speech. Measures of inhibition correlate with recognition of speech in the presence of competing talkers and in challenging environments (Dey & Sommers, 2015;

Janse, 2012; Sommers & Danielson, 1999). In this study, individual measures of working memory and inhibitory control are tested for their contribution to the various aspects of the speech recognition process.

Summary

The goal of this study was to combine behavioral and ERP methodologies to evaluate the interactions of talker accent and predictability on speech processing, and to examine any age-related changes in the context benefit. In this study, neural processing and speech recognition are compared for target words that either have high or low cloze probability based on a carrier sentence. The stimuli are produced by both native and non-native speakers of English. In order to comprehensively examine the effects of aging, context, and talker native language on speech processing, event-related potentials were measured in response to the stimuli, and listeners reported the target word after each sentence. The following outcomes were examined:

Electrophysiology:

1) Auditory object formation (P200 component)
2) Lexical access (N400 component)

Behavior:

1) Word identification accuracy
2) Word identification response time

Given the behavioral findings of increased reliance on semantic context under conditions of acoustic degradation, it was anticipated that both younger and older adults would show N400 effects of context in both the native and non-native speech conditions. Should older adults demonstrate a reduced N400 effect of context, the electrophysiologic responses would allow for a determination of whether this age effect arises from an inability to benefit from rich semantic context or an exacerbated detriment of processing non-native speech. The time-course data were expected to reveal reductions in the effects of talker over time in younger adults, consistent with the findings of Romero-Rivas et al. (2015). Evidence of rapid adaptation was expected to be delayed and/or reduced for older adults (Adank & Janse, 2010; Bieber & Gordon-Salant, 2017).

Method

Participants

The participants for this study comprised two groups of 15 listeners, including younger listeners with normal hearing (YNH) and older listeners with normal hearing (ONH). None of these listeners participated in Experiment 1. Normal hearing is defined as pure-tone thresholds of ≤25 dB HL at octave frequencies from 250-4000 Hz. Listeners reporting a history of middle ear disease or neurologic impairment were excluded from participation. Prior to testing, all listeners also completed a screening test for mild cognitive impairment (MoCA, Nasreddine et al., 2005). Listeners who did not fit the hearing-related criteria or pass the MoCA (score ≥ 26) were excluded from participation. Additionally, all listeners were required to have at least a high school education, to speak only American English as their first language, and to report no languages other than English spoken in the home before the age of 7. Listeners were also queried regarding their language history and prior exposure to non-native speech.

Stimuli and procedure

Stimuli. Stimuli for this experiment included 200 high predictability (HP) and low predictability (LP) revised Speech-In-Noise (R-SPIN; Bilger et al., 1984) sentences, recorded by two male talkers. One native speaker of English (NE) and one native speaker of Spanish (NS) were recruited from the UMD community. The recordings were made using a Shure MS48 microphone and a Marantz Professional PMD661 Handheld Solid State Recorder. Stimuli were spliced from the raw recordings using Adobe Audition 2018, and equalized for root-mean-square (RMS) amplitude using Praat (Boersma & Weenink, 2019). A 1000 Hz calibration tone that was equal in RMS level to the sentence stimuli was generated in Praat. These HP and LP R-SPIN stimuli were selected for their design in pairing monosyllabic target words within high and low predictability contexts, allowing for examination of sentential semantic context on identical target words. The sentences are phonetically balanced, and controlled for uniformity in length, with target keywords controlled for lexical frequency. Use of these stimuli also allows for comparison with the prior behavioral studies that used this corpus (Dubno et al., 2000; Pichora-Fuller, 2008; Pichora-Fuller et al., 1995; Sheldon et al., 2008).

Procedure. A total of four conditions was evaluated: High predictability, native English talker (HP, NE); low predictability, native English talker (LP, NE); high predictability, native Spanish talker (HP, NS); low predictability, native Spanish talker (LP, NS). Each listener heard 50 sentences per condition, but the HP and LP items were presented randomly within the same list, resulting in one list of 100 trials per talker. Each listener heard each target word only once, and the assignment of target words to HP/LP and NE/NS was randomized across participants. Order of talker presentation was randomized across participants and groups.

Stimulus presentation and behavioral response collection were completed using Presentation software (Neurobehavioral Systems, Berkeley CA). Stimuli were presented monaurally to the right ear at 75 dB SPL via an ER-1 insert earphone (Etymotic Research, Elk Grove Village, IL). Each trial included the following: a fixation screen to prompt the participants to listen (500 ms), auditory presentation of a sentence, and a 3-second long response window. The response screen visually presented a closed set of 6 options, including the target word and 5 foils. The participants’ task was to select the target item as quickly as possible by pressing one of six buttons. The foils were real English words that differed from the target words by one phoneme. This difference could occur on any of the phonemes, and was not consistent across foils. Participants used a keyboard to select which word was heard, allowing for collection of both response accuracy and reaction time. The key-press response also initiated the subsequent trial. Breaks were built into each list to allow time for eyeblinks and to ensure comfort. Prior to initiating the experiment, each listener completed a practice list of eight sentences from a different corpus produced by a NE talker who was not heard otherwise during the experiment. The purpose of this practice list was to familiarize the listener with the task and use of the response keyboard; listeners were given the option to repeat the familiarization list if they needed additional practice before beginning the experiment.

Following the listening experiment, all participants completed tasks from the NIH Cognitive Toolbox, including the Flanker Task and the List Sorting Working Memory Task (Weintraub et al., 2013). In the Flanker Task, which measures inhibitory control, participants are asked to respond to a target image that is flanked by two congruent or incongruent images. Flanker scores are calculated by comparing the performance on the two types of trials (congruent and incongruent). The List Sorting Working Memory Task (LSWMT) requires participants to both recall and sort a list of items that is presented both visually and auditory. LSWMT scores are calculated based on the number of correct trials. For both the Flanker and LSWMT measures, age-corrected scores were used in the analysis.

EEG recording and signal processing. EEG responses were recorded simultaneously to the behavioral task, at a 2048-Hz sampling frequency with the Biosemi Active Two system (Biosemi B.V., Netherlands) using a 34-channel cap (32 channels). Electrodes on the right and left earlobes (A1 and A2) served as reference electrodes, with additional electrodes placed above and beside the left eye to record eye movements. Event triggers were marked at the onset of the first word and the target word of each sentence. Data were analyzed offline with MNE-Python (Gramfort et al., 2014) and Eelbrain (Brodbeck et al., 2021). Responses were filtered off-line from 0.1-40 Hz and processed for analysis. Rejection of artifacts such as eyeblinks and heartbeats was completed using independent component analysis. Following artifact rejection, responses were separated into epochs of 1200 ms aligned with the time points of interest: start of the first word and start of the target word. Noisy epochs were removed from analysis, and channels containing excessive noise were interpolated. The average number of clean epochs per participant was 177/200 for the first word, and 174/200 for the target word.

Analysis

Average EEG Responses

First word. The effects of interest within the response to the first word of the sentence were the P200 and the N400. The P200 response to the first word of the sentence provides information about the effect of aging and talker language background on processing of acoustic information, while the N400 response to the first word of the sentence could be used to examine the relative ease of lexical processing for the two talkers, absent any context manipulations. In order to examine the P200 response, the average response from the Cz and Fz sensors for each individual subject was plotted, and P200 latency and amplitude were marked by hand. This strategy was used rather than analyzing a pre-determined time window due to the nature of the stimuli; as the sentence onsets were not phonemically uniform, the averaged responses were broader and less distinct than the typical P200 elicited by uniform tones or speech syllables. In addition, examination of the grand averaged waveforms confirmed the hypothesis that there would be age-related latency differences in the P200, which would necessitate a very broad analysis window. The individual P200 latencies were used to calculate individual P200 amplitudes; a window of 50 ms around each individual's peak was used to generate the average P200 amplitude across Cz and Fz. P200 latencies and amplitudes were then analyzed using two-way ANOVAs including Age Group and Talker as independent variables, with Age Group as a between-subjects variable, and Talker as a within-subjects variable. To examine responses for an N400 effect of talker and any potential interactions of talker and age group, mass univariate statistics were employed (Maris & Oostenveld, 2007). The analysis examines significance at all time bins; significant bins that occur at adjacent time points are clustered together, and then these clusters are evaluated for their likelihood of occurrence under the null hypothesis. This cluster-based nonparametric approach is recommended to control Type I error rates in electrophysiology experiments, while maintaining a conservative approach to correct for multiple comparisons (Luck & Gaspelin, 2017).

Target word. The response to the target word of the sentence was analyzed with a focus on the N400 component. A time window of 300-500 ms relative to the onset of the target word was selected for this analysis, based on the prior literature around this component's characteristics, as well as an examination of the grand mean waveforms. The mean amplitude across 300-500 ms was then analyzed using a three-way ANOVA with Age Group as a between-subjects independent variable, and Talker as a within-subjects independent variable. This analysis included only the trials in which subjects had provided a correct behavioral response (Lau et al., 2009), ensuring that the effects reflected differences in processing, rather than differences in comprehension. For the younger listeners, an average of 149 trials (77 NE) were included per subject, and for the older listeners, an average of 159 trials (80 NE) were included per subject.

Time-course patterns

Time-course analyses were conducted on all EEG amplitude measures (P200, N400 – first word, N400 – target word) as well as the behavioral measures (word identification accuracy, word identification response time). Word identification accuracy was calculated per trial, and response times were measured via keypress. Relative reaction times were calculated with the individual mean NE-HP condition RT scores serving as baseline, which was subtracted from each trial's RT for that individual. For each dataset, the raw data were plotted first in order to help determine whether a linear or non-linear analysis was more appropriate. For non-linear analyses, a growth-curve modelling approach was used, with independent orthogonal polynomial time terms evaluated in the models for contribution to model fit. All time course models were fit in R using the lme4 package (Bates, 2007) following the recommendations of Mirman (2014) and Hox et al. (2010) for forward-selection model building.

Results

Behavioral results

Word identification accuracy. Word identification scores were fitted to a generalized linear mixed effects regression model designed to examine the effects of context, talker, and age group over time. Based on visualizations of the raw data, first and second-order orthogonal polynomial time terms were created in order to independently describe the possible linear and non-linear features of the performance curve (Mirman, 2014). These terms, as well as the fixed effects of listener group (reference level = YNH), predictability (reference level = HP), talker language (reference level = Native English) and their interactions, were included sequentially in the model using forward-selection, with likelihood ratio testing used to determine whether each term significantly improved model fit following the recommendations of Hox et al. (2010). The final model selected to describe the data was $Accuracy \sim Predictability * \ Group * \ Talker + TrialNumber_{linear} + (Talker|Subject) + (Talker|Word)$. See Figure 4.1 for visualizations of the interactions of talker, predictability, and listener age.

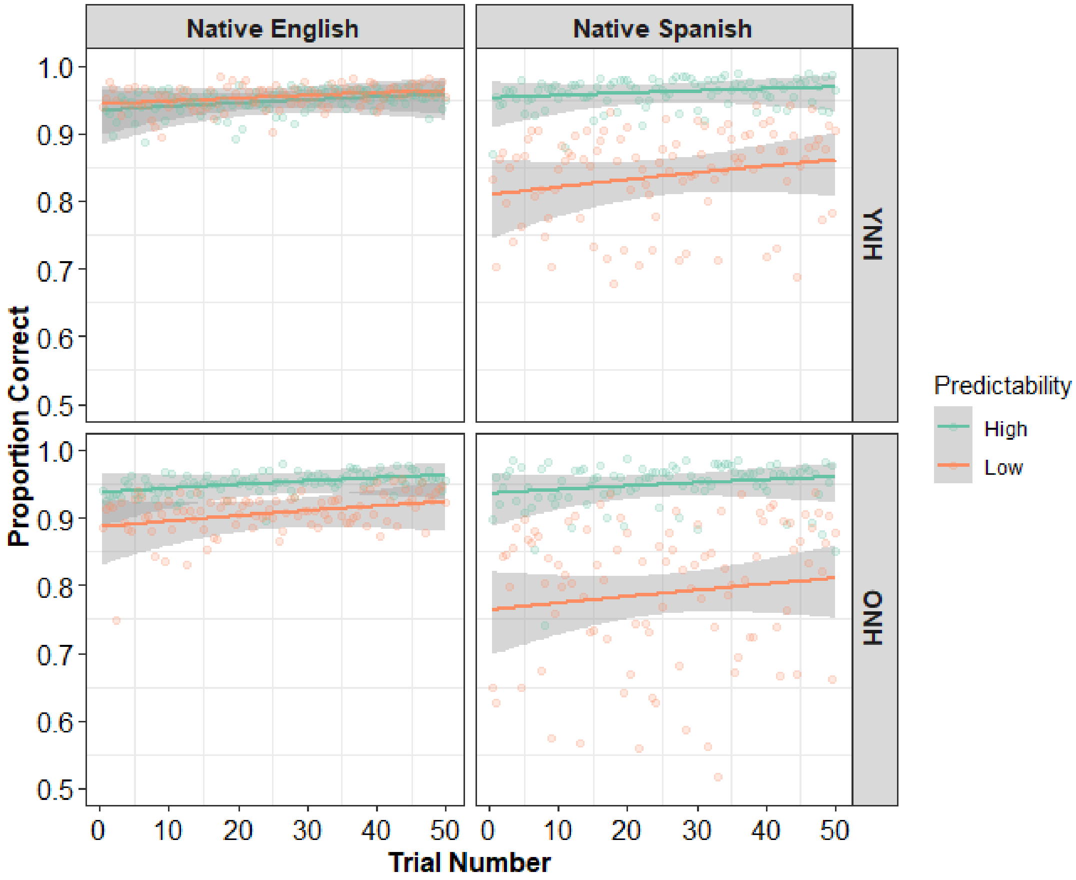


Figure 4.1. Word identification accuracy, separated by listener group, talker language background, and target predictability. Word identification performance is plotted as a function of trial number. Individual points represent group means per trial; lines reflect model predicted values with shading reflecting standard error. HP = high predictability, LP = low predictability, YNH = younger normal hearing, ONH = older normal hearing.

The full model output is included in Table 4.1.

Table 4.1. GLMER for word identification accuracy by trial. Reference levels: Group = YNH, Talker = NE, Predictability = HP. ICC = Intraclass correlation coefficient

	Word Identification Accuracy			
Predictors	*Odds Ratios*	*std. Error*	*z-value*	*p*
(Intercept)	26.53	7.45	11.68	<.001
Predictability [LP]	1.20	0.30	0.74	.46
Talker [Native Spanish]	1.59	0.49	1.50	.13
$\text{TrialNumber}_{\text{linear}}$	4.11	2.24	2.60	<.01
Group [ONH]	1.08	0.41	0.21	.83
Predictability [LP] * Talker [Native Spanish]	0.16	0.05	-5.43	<.001
Predictability [LP] * Group [ONH]	0.41	0.13	-2.71	<.01
Talker [Native Spanish] * Group [ONH]	0.62	0.24	-1.23	.22
(Predictability [LP] * Talker [Native Spanish]) * Group [ONH]	2.48	1.11	2.04	<.05
Random Effects				
σ^2	3.29			
$\tau_{00\ \text{Word}}$	0.29			
$\tau_{00\ \text{Subject}}$	0.58			
$\tau_{11\ \text{Word.TalkerNative Spanish}}$	1.18			
$\tau_{11\ \text{Subject.TalkerNative Spanish}}$	0.10			
$\rho_{01\ \text{Word}}$	-0.15			
$\rho_{01\ \text{Subject}}$	-0.93			
ICC	0.27			
N_{Subject}	30			
N_{Word}	200			
Observations	5999			
Marginal R^2 / Conditional R^2	0.095 / 0.336			

The significant three-way interaction between Predictability, Group, and Talker resulted from a larger predictability effect for the native Spanish talker than for the native English talker, which was amplified for the ONH listeners as compared to the YNH listeners ($\beta = 0.91$, SE = 0.45, $z = 2.04$, $p < .05$). Further examination of the interaction reveals that the ONH listeners showed predictability effects for both the native English ($\beta = 0.71$, SE = 0.22, *z-ratio* =

3.2, $p < .01$) and native Spanish ($\beta = 1.62$, SE = 0.22, $z = 7.43$, $p < .001$) talkers, whereas the YNH listeners did not show a predictability effect for the native English talker ($\beta = -0.19$, SE = 0.25, $z = -0.74$, $p = .46$). Within the high predictability condition, neither the YNH ($\beta = -0.46$, SE = 0.31, $z = -1.51$, $p = .13$) nor the ONH ($\beta = 0.01$, SE = 0.3, $z = 0.04$, $p = .97$) listeners showed an effect of talker, suggesting that the presence of non-native talker accent alone did not significantly reduce performance, when supportive semantic context was available.

The significant main effect of Trial number indicated that performance increased significantly across trials ($\beta = 1.41$, SE = 0.54, $z = 2.6$, $p < .01$). The quadratic time term was found not to contribute significantly to model fit ($p>0.5$), and the linear time term was found not to interact significantly with any of the other fixed predictors ($p>.05$, all comparisons), suggesting that the performance increases were similar across all conditions. Listeners' performance on the cognitive tasks was also tested for contribution to model fit, but none were found to significantly improve the model ($p>.05$, all comparisons).

In summary, older adult listeners had lower word identification accuracy for target words in low-predictability contexts, especially for stimuli spoken by a Spanish-accented talker. Younger adults also showed this predictability effect for the Spanish-accented stimuli, but not for unaccented stimuli. All listeners improved their word identification accuracy over time, regardless of talker or predictability.

Reaction times. Relative reaction times were fitted to a linear mixed effects regression with similar procedures to those described above for the accuracy analysis. The final model selected to describe the relative reaction time data was: $RT_{rel} \sim Predictability * \ Talker + (TrialNumber_{linear} * Group\ ) + (1|Subject) + (1|Word)$. Inclusion of random slopes was evaluated, but found to result in singular fits, indicating overfitting; random slopes were excluded from the final model. The relative reaction times are visualized in Figure 4.2.

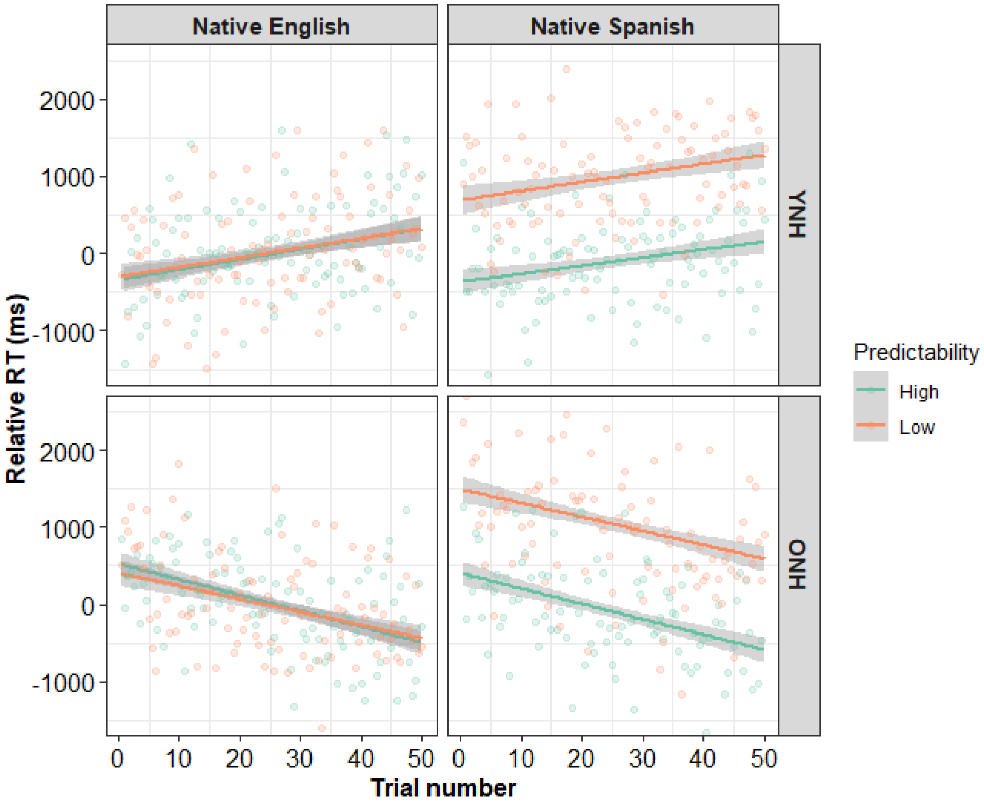


Figure 4.2. Fitted values for relative reaction time (RT), separated by listener group, talker language background, and target predictability. Relative RT is plotted as a function of trial number. Relative RT was calculated by subtracting the individual's mean RT in the Native English HP condition from their RT in each trial. Individual points represent group means per trial; lines reflect model predicted values with shading reflecting standard error. HP = high predictability, LP = low predictability, YNH = younger normal hearing, ONH = older normal hearing.

The significant interaction of Predictability and Talker reflects the presence of a context effect for the NS talker, but not for the NE talker (β = 1064.26, SE = 258.64, t = 4.12, p < .001). Relative reaction times are slower for the NS talker in the low predictability sentences, consistent with increased effort associated with the word identification process in these conditions. The interaction of trial number and listener group shows that, regardless of talker, the relative RT increased over time for the YNH listeners, but decreased over time for the ONH listeners. The effect of trial number on RT is significantly more negative for the ONH listeners than for the YNH listeners (β = -31.37, SE = 8.79, z = -3.57, p < .001). Listener performance on the cognitive tasks was also tested for contribution to model fit, but none were found to significantly improve the model (p>.05, all comparisons). The full model summary is found in Table 4.2.

Table 4.2. LMER for relative reaction times by trial. Reference levels: Group = YNH, Talker = NE, Predictability = HP. ICC = Intraclass correlation coefficient.

	Relative Reaction Time			
Predictors	*Estimate*	*std. Error*	*t-value*	*p*
(Intercept)	-298.96	293.76	-1.02	.31
Predictability [LP]	69.67	184.84	0.38	.71
Talker [Native Spanish]	-97.21	176.25	-0.55	.58
Group [ONH]	818.59	363.80	2.25	<.05
TrialNumber$_{linear}$	11.03	6.39	1.73	.08
Predictability [LP] * Talker [Native Spanish]	1064.26	258.63	4.11	<.001
Group [ONH] * TrialNumber$_{linear}$	-31.37	8.78	-3.57	<.001
Random Effects				
σ^2	21811983.49			
$\tau_{00\ \mathrm{Word}}$	1348082.24			
$\tau_{00\ \mathrm{Subject}}$	499372.41			
ICC	0.08			
N_{Subject}	30			
N_{Word}	200			

Observations	5458
Marginal R^2 / Conditional R^2	0.010 / 0.088

Thus, relative reaction times were slower for low predictability than high predictability stimuli, but only for the stimuli produced by the Native Spanish talker. Younger listeners showed steadily increasing relative RTs over time, while ONH listeners' RTs decreased with additional listening in each condition.

Average EEG responses

First word: P200. The response to the first word of the sentence is displayed in Figure 4.3, with separate curves (and colors) depicted for each age group and talker.

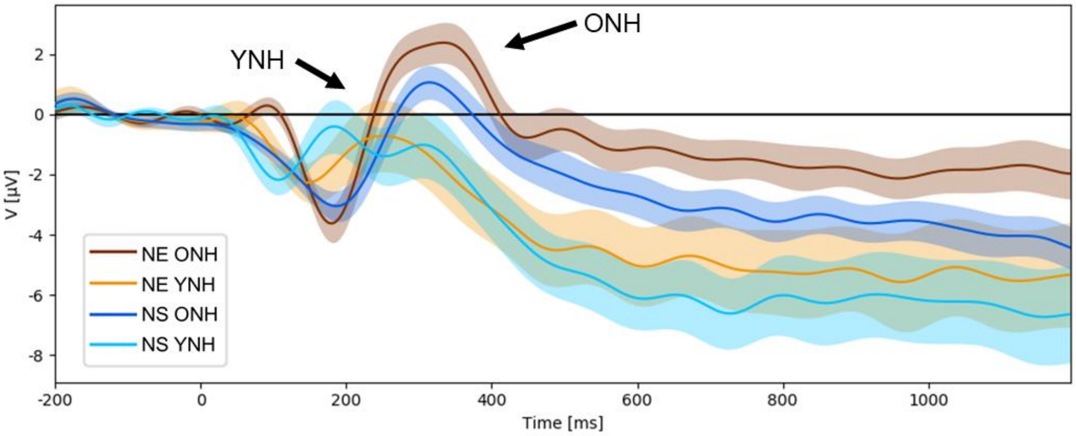


Figure 4.3. Averaged ERP waveforms in response to the first word of the sentence, separated by listener group and talker language. The time of 0 ms corresponds to the onset of the first word. The shaded region denotes standard error. Responses are averaged over the Cz and Fz sensors and filtered from 0.1-20 Hz for visualization purposes only. NE = Native English, NS = Native Spanish, YNH = younger normal hearing, ONH = older normal hearing. Arrows indicate the regions of the P200 peaks for the two listener groups.

The analysis of P200 latency indicates a main effect of group (F(1, 56) = 13.63, *p*<.001), with the older adults showing later P200 latency than younger adults. P200 latencies are displayed in Figure 4.4.

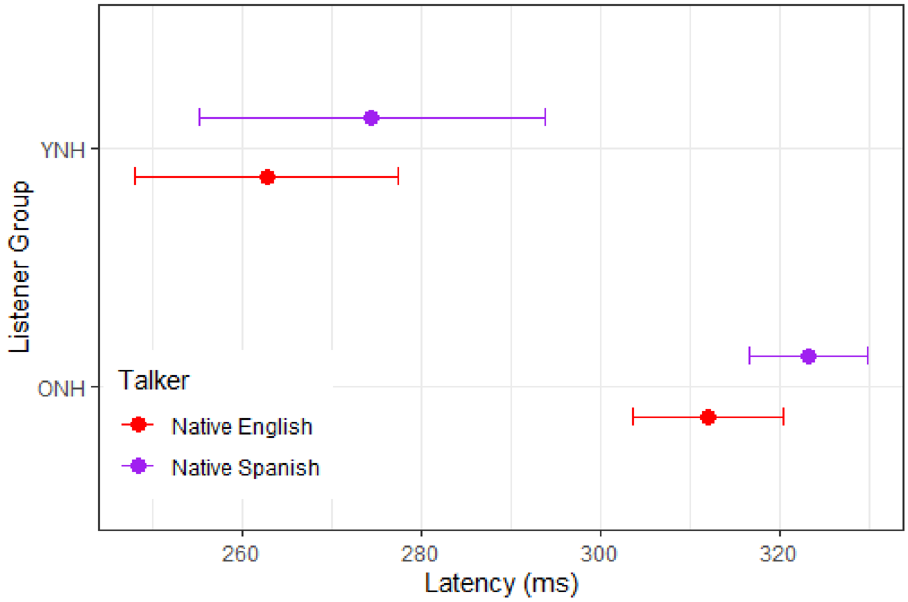


Figure 4.4. Latencies of the P200 component in response to the first word of the sentence, separated by listener group and talker type. Error bars reflect standard error of the mean. YNH = younger normal hearing, ONH = older normal hearing.

There is no main effect of talker or interaction between talker and age group on P200 latency. The analysis of P200 amplitude does show an interaction of age group and talker (F(1, 56) = 5.69, *p*<.05), with younger adults showing no significant differences in amplitude between the response to the NE talker and the NS talker. Older adults, however, showed a decrease in P200 amplitude in the response to the NS talker as compared to the NE talker.

In summary, listener age had an effect on both the latency and amplitude of the P200 component, with older adults showing delayed latencies and exaggerated amplitudes. In addition, older adults' responses were influenced by talker, with P200 amplitude reduced for the native Spanish talker. This talker effect was not seen for the younger adults.

First word: N400. The cluster-based analysis revealed a main effect of talker beginning at 459 ms and lasting for the duration of the analysis window (459-1200 ms, $p<.01$), and a main effect of group beginning at 278 ms and lasting for the duration of the analysis window (278-1200 ms, $p<.01$). The distribution of the talker effect (See the bottom right topography in Figure 4.5) is consistent with an N400-like effect, which is typically distributed over the central cortical regions. This indicates decreased ease of lexical access for the NS talker as compared to the NE talker; the effect did not interact with listener age.

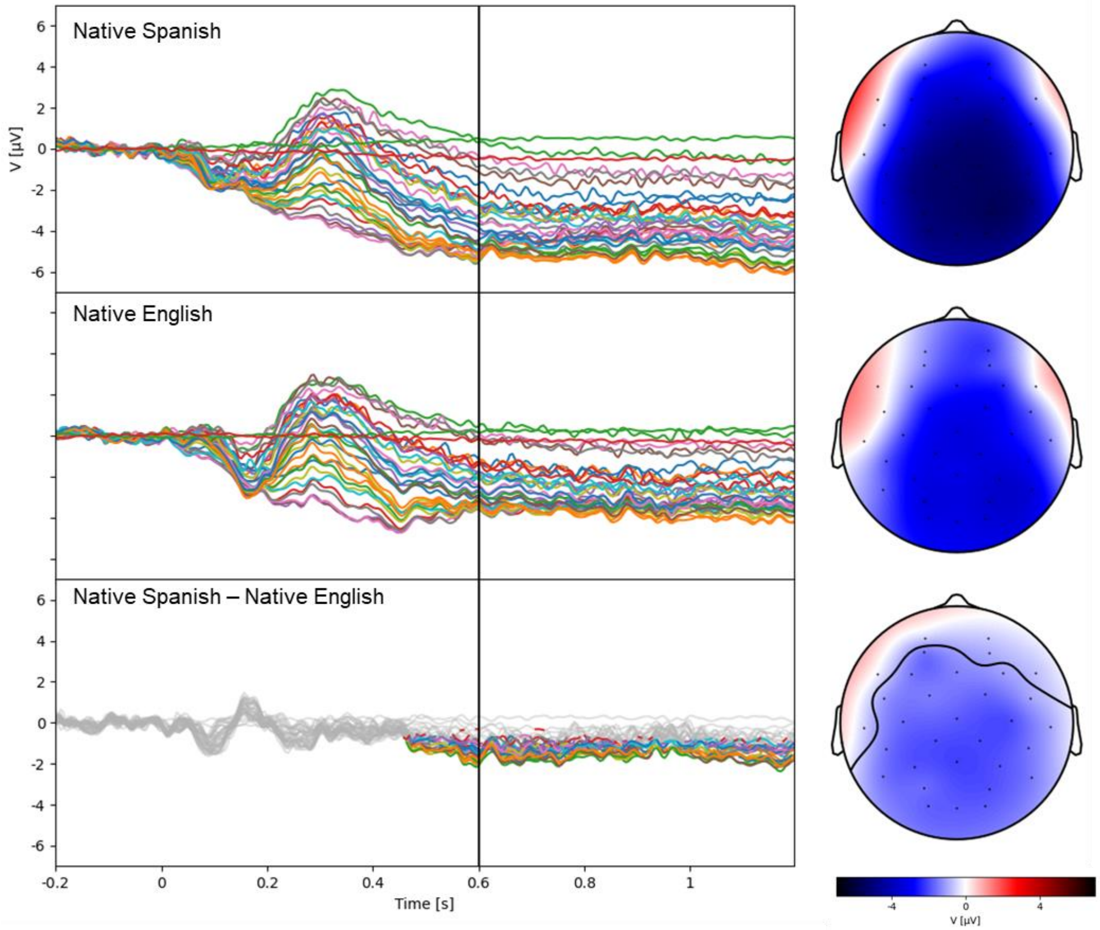


Figure 4.5. Visualization of the talker effect in response to the first word. Top = Native Spanish (NS), middle = Native English (NEs), bottom = NS – NE. Responses are averaged across listener group and predictability. The solid vertical line represents time at 600 ms; the topographies for the talker effect at 600 ms are shown on the right. Channels with significant difference are bounded in the bottom topography map, indicating a centro-parietal distribution.

This finding is consistent with that seen by Romero-Rivas et al. (2015), who found that young adult listeners had larger N400 amplitudes in response to words spoken by non-native talkers compared to native talkers, independent of any context or predictability manipulations.

Target word: N400.

In Figure 4.6, average responses to the target word are displayed for both talker/listener group combinations, as well as difference waves comparing the LP and HP conditions.

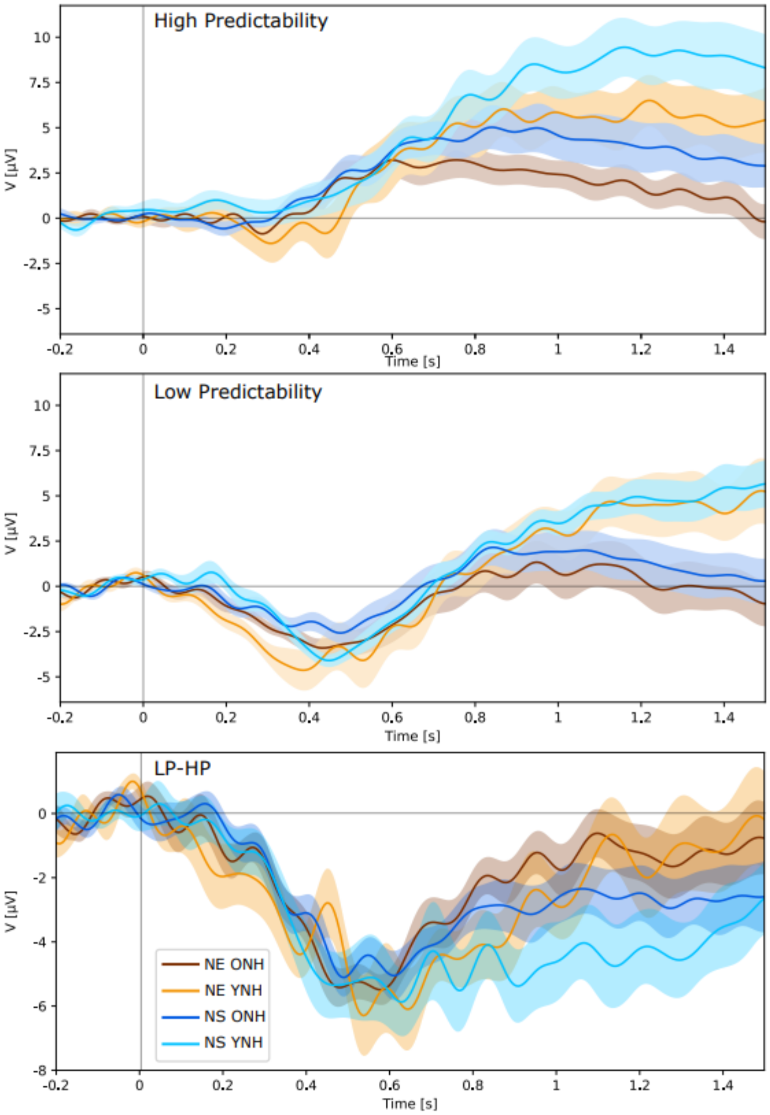


Figure 4.6. Averaged ERP waveforms in response to the target word of the sentence, separated by listener group, talker language, and target predictability. Responses to target words in high-predictability (HP) and low-predictability (LP) sentences are shown in the top and middle panels, respectively. The bottom panel reflects the predictability difference wave (LP-HP). 0 ms corresponds to the onset of the target word. The shaded region

denotes standard error. Only trials with correct behavioral responses are included. Responses are averaged over the Cz and Pz sensors and filtered from 0.1-20 Hz for visualization purposes only. NE = Native English, NS = Native Spanish, YNH = younger normal hearing, ONH = older normal hearing

A mass-variate repeated-measures ANOVA was performed for the average amplitude between 300-500 ms across all sensors, assessing the within-subject effects of predictability and talker language background, and the between-subject effect of listener group. The analysis revealed a significant effect of predictability (p<.001). The predictability effect, as expected, showed a greater magnitude of deflection (i.e. lower absolute amplitude) for the LP sentences. The central distribution of the response is consistent with the classic N400 response; see Figure 4.7 for visualization of the effect and its distribution.

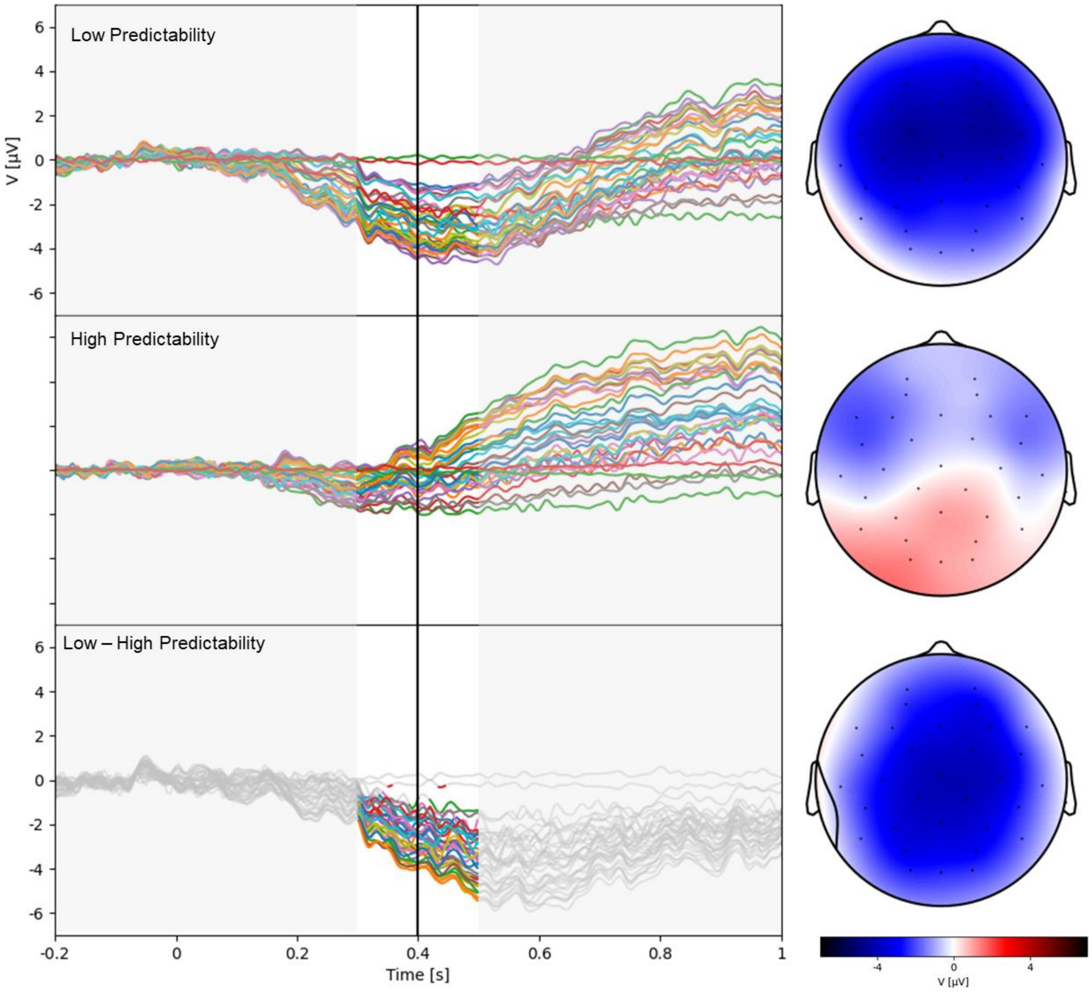


Figure 4.7. Visualization of the predictability effect in response to the target word. Top = Low predictability (LP), middle = High predictability (HP), bottom = LP – HP. Responses are averaged across listener group and talker; only trials with correct behavioral responses are included. The solid vertical line represents time at 400 ms; the topographies for the predictability effect at 400 ms are shown on the right. Channels with significant difference are bounded in the bottom topography map; nearly all channels are significant in this analysis. The highlighted waveform region of 300-500 ms reflects the time boundaries of the analysis window.

There was no main effect of group, nor did group interact with any other predictor variable. However, a significant effect of talker (*p*<.05) was observed, indicating that amplitudes were more negative for the NE talker as compared to the NS talker; see Figure 4.8 for visualization of this effect. However, talker and predictability did not interact in this analysis.

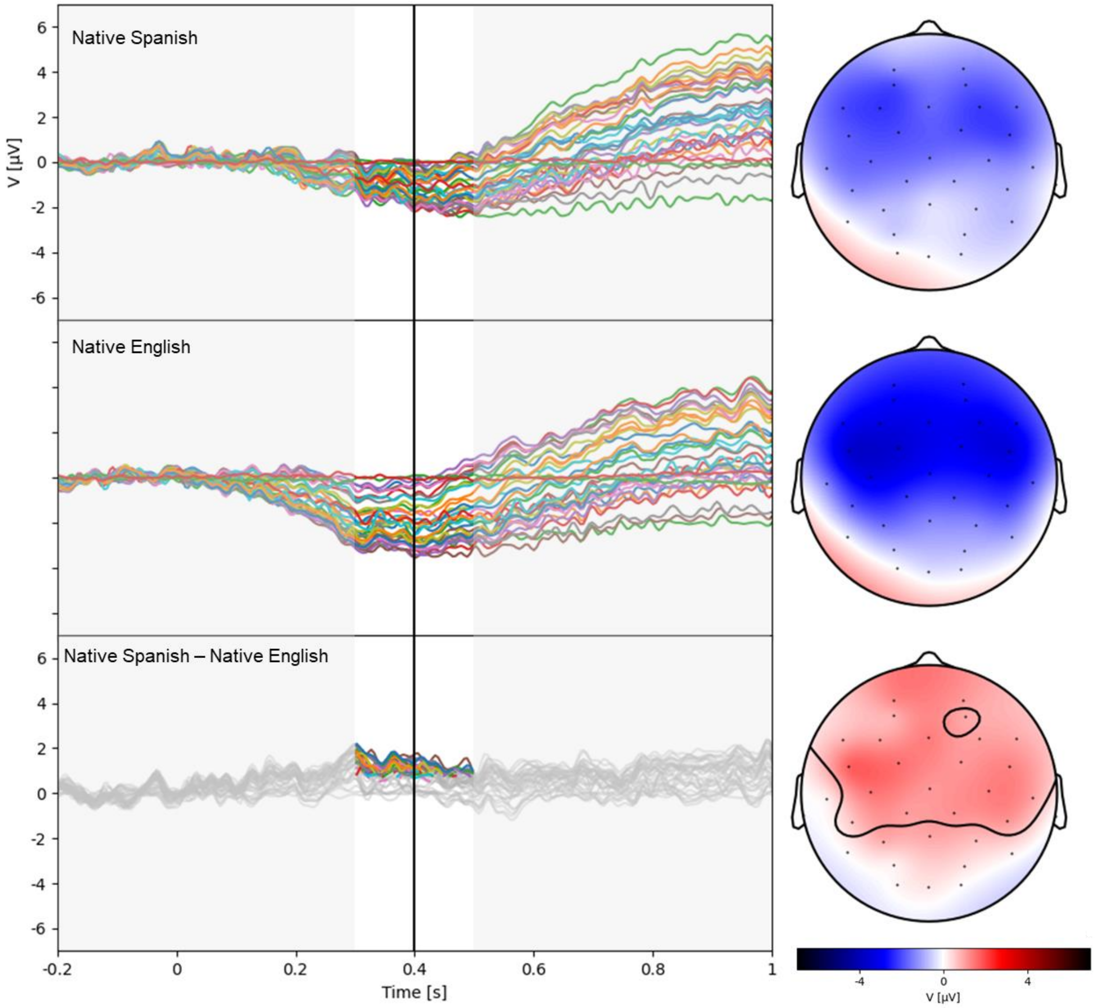


Figure 4.8. Visualization of the talker effect in response to the target word. Top = Native Spanish (NS), middle = Native English (NE), bottom = NS – NE. Responses are averaged across listener group and predictability. The solid vertical line represents time at 400 ms; the topographies for the talker effect at 400 ms are shown on the right. Channels with significant difference are bounded in the bottom topography map, indicated a centro-frontal distribution of the effect. The highlighted waveform region of 300-500 reflects the time boundaries of the analysis window.

The talker effect seen in this 300-500 ms time range reflects a reversal of the talker effect seen in response to the first word of the sentence, where the response to the NS talker was more negative than that for the NE talker. The centro-frontal distribution of the talker effect within the 300-500 ms time range also differs from the distribution of the talker effect in response to the first word of the sentence, which had a more centro-parietal distribution. Electrode sensor was also evaluated as a predictor to test for differences in distribution of the N400 response across listener groups or talker types, but no significant interactions were found (*p*>.05, all comparisons).

Time-course patterns

First word: P200 and N400. The effect of trial number on P200 amplitude was evaluated by iteratively building a linear mixed effects regression model with the individual P200 amplitudes across the Cz and Fz channels included as the dependent variable. Trial number did not have a significant impact on P200 amplitude and did not interact with any of the fixed predictors of talker or listener group (*p*>.05, all comparisons). Similarly, trial number did not predict the amplitude of the N400 response to the first word and did not interact with group or talker. Thus, the group and talker-related effects on these responses described above (Average EEG Responses section) remained constant.

These findings indicate that the P200 and N400 component amplitudes in response to the first word of the sentence did not change significantly over the course of any listening condition. The stability of the P200, indicating no change to the processing of acoustic input, is consistent with prior work showing that training-induced changes to P200 do not occur within-session

(Atienza et al., 2002; Romero-Rivas et al., 2015). The stability across trials of the N400 response to the first word suggests that the process of lexical access required greater processing resources for the NS than for the NE talker, even with additional listening experience.

Target word: N400. The size of the N400 component in response to the target word of the sentence was examined over the course of the experiment by averaging the amplitude across the Cz and Pz sensors between 300-500 ms for each trial. A linear mixed effects regression was fitted to these mean amplitudes to analyze the effects of predictability (reference = HP), talker (reference = Native English), and listener group (reference = YNH) over the course of the listening conditions. The model was built using forward selection following the recommendations of Hox et al. (2010), with likelihood ratio testing used to confirm the inclusion or removal of each term within the model. The final model selected to describe the data was: $MeanN400 \sim Predictability + Talker + (TrialNumber_{linear} * Group) + (Talker * WorkingMemory) + (Predictability|Subject) + (Predictability|Word)$. These effects are visualized in Figure 4.9.

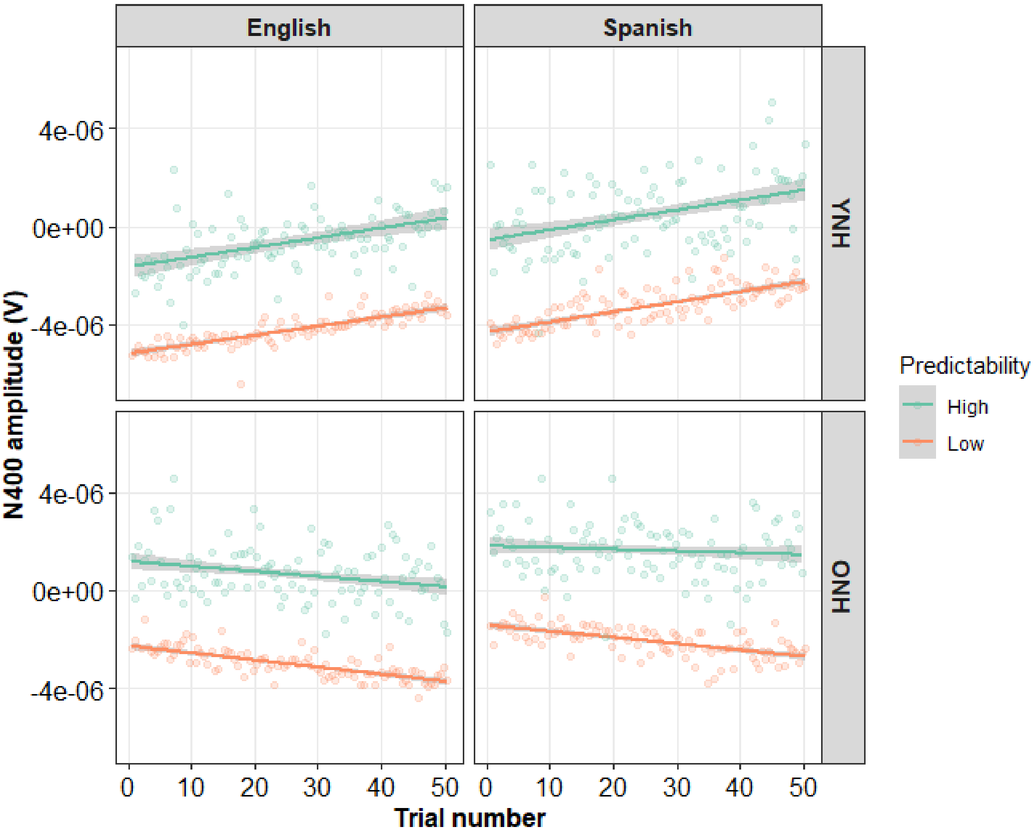


Figure 4.9. Fitted values for mean N400 amplitude in response to the target word, separated by listener group, talker language background, and target predictability. Amplitude is plotted as a function of trial number; a more negative amplitude indicates a larger magnitude N400 component. Individual points represent group means per trial; lines reflect model predicted values with shading reflecting standard error. HP = high predictability, LP = low predictability, YNH = younger normal hearing, ONH = older normal hearing

The main effects of predictability (β = -3.56e-06, SE = 4.7e-07, t = -7.57, p<.001) and talker (β = 8.47e-07, SE = 3.01e-07, t = 2.81, p<.01) are consistent with the findings of the average amplitude model (see above), as is the lack of their interaction. The significant interaction of trial number and group (β = -6.4e-08, SE = 2.07e-08, t = -3.1, p<.01) indicates that the two listener groups showed different patterns of N400 component amplitude change over time. The mean N400 amplitude decreased in magnitude over the course of trials for the YNH listeners, while the ONH listeners showed the opposite pattern. The significant interaction between talker and working memory scores (collapsed across listener groups) is visualized in Figure 4.10, and suggests that there was a stronger relationship between working memory scores and N400 component amplitude for target words produced by the NE talker than the NS talker (β = 8.72e-07, SE = 3.04e-07, t = 2.87, p<.01).

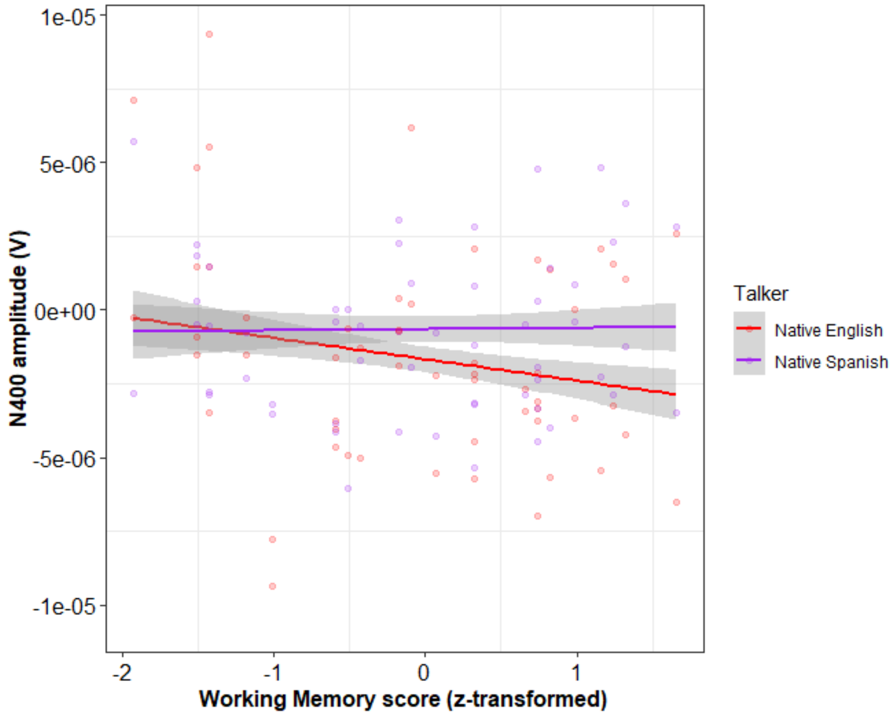


Figure 4.10. Relationship between mean N400 amplitude in response to the target word and scores on the List Sorting Working Memory Task (LSWMT). LSWMT scores were age-corrected and z-transformed for analysis. Separate regression lines are plotted for each talker. Data are collapsed by group and target predictability. Points represent individual data; regression lines reflect model predicted values with shading reflecting standard error.

Specifically, individuals with better working memory performance had N400 components with larger magnitudes, or more negative absolute amplitudes. Full details of the model can be found in Table 4.3.

Table 4.3. Target word N400 component by trial. Reference levels: Group = YNH, Talker = NE, Predictability = HP. ICC = Intraclass correlation coefficient.

	Mean N400 amplitude			
Predictors	*Estimate*	*std. Error*	*t-value*	*p*
(Intercept)	-1.18e-06	7.51e-07	-1.57	.12
Predictability [LP]	-3.55e-06	4.7e-07	-7.57	<.001
Talker [Native Spanish]	8.47e-07	3.01e-07	2.81	<.01
Group [ONH]	2.55e-06	8.16e-07	3.13	<.01
$\text{TrialNumber}_{\text{linear}}$	3.24e-08	1.49e-08	2.18	<.05
ListSort_z	-1.04e-06	3.45e-07	-3.03	<.01
Group [ONH] * $\text{TrialNumber}_{\text{linear}}$	-6.4e-08	2.07e-08	-3.10	<.01
Talker [Native Spanish] * ListSort_z	8.72e-07	3.04e-07	2.87	<.01
Random Effects				
σ^2	9.73e-11			
$\tau_{00\ \text{word}}$	3.53e-12			
$\tau_{00\ \text{subject}}$	6.77e-12			
$\tau_{11\ \text{word.predLP}}$	6.32e-12			
$\tau_{11\ \text{subject.predLP}}$	2.7e-12			
$\rho_{01\ \text{word}}$	-0.94			
$\rho_{01\ \text{subject}}$	-0.95			
ICC	0.06			
N_{subject}	28			
N_{word}	200			
Observations	4375			
Marginal R^2 / Conditional R^2	0.042 / 0.102			

In sum, YNH listeners showed decreasing N400 components for both talkers and predictability conditions with additional listening experience, while older adults did not. Working memory capacity predicted the overall magnitude of the N400 component for the NE talker more strongly than for the NS talker, regardless of listener group.

Discussion

Effects of talker on acoustic processing

The EEG measures in this study allow for examination of several points in the speech recognition process. The earliest stage of auditory processing examined in this study, acoustic feature extraction (as indexed by the P200 component), was impacted by non-native accent for older but not younger listeners. Older adults showed a reduction in the amplitude of the P200 response for the NS talker compared to the NE talker, suggesting that the older adults had more difficulty with auditory object formation for the NS as compared to the NE talker. This was seen despite the overall level of overcompensation shown in the older adults' P200 response (i.e. exaggerated P200 amplitude) relative to the younger listeners. The P200 component was also delayed in latency and increased in amplitude for older adults. These findings of delayed latency and exaggerated amplitude for the P200 are expected given prior literature about aging effects in the auditory cortex (Roque et al., 2019; Tremblay et al., 2003). The effect of talker on P200 amplitude remained constant over the course of the experiment, suggesting that any learning of talker or accent was not evident in changes to the P200 response within this experiment. Prior literature has shown that the P200 may be a marker of auditory training, and that changes typically do not occur within the training session itself, but emerge across sessions (Atienza et al., 2002; Rossi et al., 2013; Tremblay et al., 2001).

Effects of talker and context on lexical processing

Further examination of the response to the first word of the sentence provides information about the ease of lexical access for the two talkers included in this study, before any contextual information becomes available. The main effect of talker beginning at 458 ms showed a central distribution, consistent with an N400 effect. Greater N400 component magnitude in the NS condition is interpreted to reflect an increased difficulty in mapping lexical meaning onto NS speech, independent of any manipulations of predictability, and is consistent with the N400 effect of talker seen in prior literature involving non-native speakers (Goslin et al., 2012; Romero-Rivas et al., 2015). The main effect of talker remained constant over the course of the experiment; even after listening to 100 sentences, the process of mapping lexical meaning to acoustic input required similar processing resources regardless of talker. This finding is inconsistent with that of Romero-Rivas et al. (2015), who found a reduction in the magnitude of an N400 effect of talker across experimental blocks. One possible explanation for this discrepancy is that the prior study utilized eight accented talkers from different language backgrounds, while the present study used just one NS talker. Speech recognition is typically more challenging when listening to multiple talkers vs a single talker (Mullennix et al., 1989; Sommers et al., 1994); perhaps the N400 effect of talker in the prior study was larger in magnitude than that of the present study, allowing more room for reduction over time.

The response to the final target word of the sentence allows an examination of the influence of sentential context on the questions of lexical access for NE and NS speech. Both younger and older adults showed the expected predictability effect, which is interpreted to reflect a greater allocation of processing resources when speech is lacking in semantic context, due to the increased difficulty of lexical access for these items. Additionally, both younger and older adults showed an effect of talker, where the response to the NE talker had a greater negativity than the response to the NS talker. This effect was expected to occur in reverse: a greater negativity for the NS talker would have been interpreted as an increased difficulty of lexical access for NS speech, as was seen in response to the first word of the sentence. However, this reverse effect may have occurred as a side effect of the procedure employed to create equivalent baselines prior to the analysis time periods. As evident in the response to the first word of the sentence, the NS speech evoked an overall more negative response and a greater draw on processing resources at the outset of the sentence. By the end of the sentence, when the target word was heard, this talker effect may suggest that listeners had less spare processing resources available for the NS speech than the NE speech. This is reflected in the talker effect seen in response to the target word, in which NE speech elicited a greater negativity with centro-frontal distribution, indicating greater recruitment of processing resources than for the NS speech.

Working memory

Working memory capacity was found to predict N400 component amplitude in response to the target word of the sentence, with a stronger relationship between LSWMT scores and N400 amplitude for the NE speech. Working memory was not found to be predictive of the other ERP measures, nor the behavioral word identification measures. This finding extends prior literature documenting a relationship between working memory and a context-elicited N400 effect in response to visually presented stimuli (Federmeier & Kutas, 2005; Van Petten et al., 1997). Individuals with higher working memory scores showed greater N400 component amplitudes in response to NE speech. One possible explanation for this relationship is that these listeners had a greater ability to retain and manipulate the information contained in the sentences leading up to the target word, and thus greater processing resources were utilized in mapping meaning to the target word for these listeners. The absence of this relationship for the NS speech may relate to the overall increased demand of processing more challenging speech. If listeners operate with a finite processing capacity (Kahneman, 1973), listeners may not have been able to draw on working memory resources in order to aid in the processing of the non-native speech.

Age, talker language, and predictability

When examining the average responses, the effects of predictability and talker on N400 component amplitude did not interact with each other, which was unexpected. The average response to the target word also did not show any age effects, nor interactions with age. This lack of any age effects contrasts with prior findings of an age-related reduction in the N400 effect (Federmeier et al., 2010; Federmeier & Kutas, 2005; Payne & Federmeier, 2018; Wlotko et al., 2012). Several factors could contribute to this finding. One potential discrepancy between this study and the prior ERP literatures relates to hearing thresholds. Many of the ERP studies documenting an aging detriment did not measure or report pure-tone thresholds for their listener groups, which creates a potential confound between effects due to aging alone vs age-related hearing loss. Age-related hearing loss is known to result in detriments in auditory processing above and beyond age effects alone (Anderson, Parbery-Clark, et al., 2013; Tremblay et al., 2003). In addition, the paradigm used in this study may have resulted in an increased engagement of on-line attention compared to tasks employed in prior literature. In this study, participants were asked to complete a word recognition task following each sentence, whereas previous studies employed tasks such as passive listening (Romero-Rivas et al., 2015), congruency judgements (Federmeier et al., 2003), or delayed recall (Federmeier & Kutas, 2005). The on-line word recognition task may have caused participants to devote relatively higher attention to the stimuli in this experiment, contributing to an improved ease of processing for the degraded signal (Wild et al., 2012) and eliminating the expected interaction.

The effects of predictability, age, and talker accent emerged in various cortical measures of the speech perception process, and can be considered in conjunction with an overall three-way-interaction present in the behavioral word identification scores. The effect of age on word identification accuracy may result more from age-related differences in the earlier processing of acoustic features (P200 response to first word), as the higher-level predictive processing capabilities that are indexed by the context-elicited N400 do not seem to be reduced in the older normal-hearing adults, on average. However, considering the effects of age on the magnitude of the context-elicited N400 component over the course of the experiment, younger adults show a steady decrease in N400 component magnitude with additional listening time for most conditions, while older adults do not. This stands in contrast to the identification accuracy data, where the older adults do show increases in word identification performance over the course of trials. Additionally, the reaction time data show that the older adults' relative reaction times decrease over the course of trials for all conditions except NS-LP, suggesting that their behavioral responses generally become less effortful over time.

Together, these results suggest that, while older adults may be able to improve their behavioral performance within the course of the experiment, the improvements likely do not stem from an increased ability to utilize predictive processing (N400 response to target word). This finding aligns with those of Federmeier et al. (2005), who found that age-related reductions in the visually evoked N400 effect were due to an inability to utilize rich contextual information, rather than an exaggerated susceptibility to low-context stimuli. In addition, the older adults do not show evidence of an increased ability to

extract acoustic or lexical information from the non-native speech input, independent of contextual manipulations (EEG responses to first word).
Of course, some caution should be applied when directly comparing these data, as the cortical responses to the target word are only analyzed for trials where the behavioral response was correct, and the word identification accuracy measure considers all trials in each condition. Thus, the aim is not to directly correlate the two forms of measures, but rather examine both commonly used outcome measures in the same subjects to understand the patterns displayed by older and younger listeners.

One possible factor contributing to the older adults' improvements in behavioral performance over the course of the experiment could be age-related differences in the time course of task familiarization. In the current protocol, all listeners completed a practice round before beginning the experiment. However, this practice round only consisted of 8 trials, which may not have been sufficient for the older adult listeners to completely acclimate to the task, contributing to overall lower starting performance.

Conclusion

In conclusion, this study found that lexical access for speech produced by a non-native talker required greater processing resources than for speech produced by a native talker, regardless of listener age. Older adults did not show reductions in their ability to use context for lexical processing, as indexed by the predictability-elicited N400 effect, and talker native language did not influence the magnitude of this predictability effect. Regardless of context level, younger adults appeared to show increased ease of lexical access for both NE and NS speech with additional listening experience, while older adults did not. In contrast, older adults showed decreases in relative reaction time over the course of the experiments, while younger adults did not. However, all listeners showed improvements in word identification over time, which did not differ across talker or predictability conditions, despite overall poorer performance with the low-predictability NS speech. Together, these results expand the prior literature regarding aging and use of context in speech recognition, and suggest that improvements seen in behavioral measures of speech recognition in older adults do not appear to result directly from improvements in predictive processing.

Chapter 5: Discussion

Introduction

The overarching goal of this project was to examine the intrinsic and extrinsic factors that contribute to a listener’s ability to adapt to non-native speech. The primary intrinsic factors that were probed in this project were aging, age-related hearing loss, and individual cognitive capacity. The extrinsic factors that were examined included talker accent, stimulus variability and semantic context, variables that affect the bottom-up and top-down processing of speech, respectively. The results of the three experiments presented here can be viewed together in order to gain some insights about these questions.

Aging and recognition of non-native English speech

On the whole, the experiments presented in this project indicate that aging alone does not impede recognition of non-native speech. In the majority of the experimental conditions, older adults with normal hearing sensitivity (ONH) performed similarly to younger adults with normal hearing sensitivity (YNH) when listening to non-native English (NNE) talkers. In Experiment 1, listeners completed a number of listening conditions with various configurations of non-native talkers in the presence of multitalker babble. In all but one of the NNE talker conditions, overall speech recognition levels were similar between YNH and ONH listeners, who listened at identical signal-to-noise ratios (SNRs). In Experiment 3, listeners heard a single native Spanish talker produce English words in quiet. Interestingly, there was an interaction of talker and age group when examining the P200, an event-related potential that indexes cortical processing of the acoustic features of the speech stimulus. Younger adults showed no difference in the amplitude of this auditory ERP, but older adults showed a reduction in P200 amplitude in response to Spanish-accented English speech, suggesting that this altered acoustic information was more challenging to process. However, when the words were presented in a context-rich carrier phrase, there were no significant differences in behavioral word identification accuracy between younger and older listeners.

Collectively, these findings indicate that non-native accent, which is a naturalistic acoustic alteration to the auditory signal, does not impede speech recognition in older adult listeners when tested with simple, semantically rich sentences. This was true even in the presence of background noise, suggesting that older adults with intact hearing sensitivity were not negatively affected by changes to the bottom-up signal features more than the younger adults. Aging effects did emerge when the top-down features were manipulated; this will be discussed in more detail below.

Aging and adaptation to non-native English speech

Aging did not appear to have a detrimental effect on the ability to rapidly adapt to non-native English speech. Rapid adaptation was evaluated for a number of different listening conditions in this project. In Experiment 1, listeners adapted to non-native talkers in single and multiple talker conditions. Differences in overall level of performance were observed across the talker conditions, but when the two normal-hearing listener groups were compared, similar patterns of adaptation were found. That is, YNH and ONH listeners displayed the same time course of performance changes when listening to a set of 30 sentences. In Experiment 3, the word identification measures were consistent with the results of Experiment 1: older and younger adults showed similar rates of adaptation to both NE and NNE speech.

Additionally, there were no differences between older and younger listeners in terms of magnitude of adaptation, or generalization of learning to an unfamiliar talker. It was anticipated that younger and older normal-hearing adults would show a similar magnitude of adaptation, as this has been documented in some prior studies of short-term learning for non-native speech in younger and older normal-hearing adults (Adank & Janse, 2010; Gordon-Salant, Yeni-Komshian, Fitzgibbons, et al., 2010). Some differences in rate or time course of adaptation were expected, however. Age-related differences in patterns of learning have been observed in studies utilizing various forms of challenging speech, including time-compressed speech (Peelle & Wingfield, 2005), ambiguous phonemes (Scharenborg & Janse, 2013), speech-in-noise (Karawani et al., 2016) and non-native speech (Adank and Janse, 2010). Other studies have found no differences in patterns of adaptation between younger and older normal-hearing listeners (Erb & Obleser, 2013; Neger et al., 2014); the present results support these findings.

Rapid adaptation was tested in younger and older listeners with an expanded set of measurements in Experiment 3, including both behavior and electrophysiology. When measures besides word identification accuracy (EEG, relative reaction times) were examined, there were some differences between younger and older normal-hearing listeners. It should be emphasized that the various methods used in Experiment 3 do not measure the same processes, rather these measures can be used to provide insight into different components of the speech recognition process. For example, reaction time measures for the word identification process, which were only analyzed for correct responses, can provide information about how much effort was expended to generate accurate word identification (Clarke & Garrett, 2004; Munro & Derwing, 1995).

In Experiment 3, ONH listeners showed consistent decreases in relative reaction time over the course of trials, suggesting that successful word identification became less effortful with additional listening time for these listeners. One possible explanation is that the older adults experienced rapid perceptual adaptation to the talkers, which was evident in both identification accuracy and response time. On average, older adults' responses became more accurate over time, and they were able to produce accurate responses more quickly. However, the same pattern was not seen in the younger adult group. In fact, their relative reaction times for correct responses grew longer over the course of the experimental conditions, even as their accuracy improved. One possible explanation for the increased reaction times in the younger adults could be an increase in fatigue as the experiment progressed.

Another method for examining rapid adaptation within the listening conditions in Experiment 3 was measuring change in ERP amplitudes across trials. Several ERP measures were tracked, including indices of acoustic and lexical processing. One of these measures did show changes within the course of the experiment: N400 component amplitude in response to the target word. In this study, the N400 component reflects the processing resources recruited during the process of mapping lexico-semantic meaning to the target acoustic signal. N400 component analyses were also only conducted on trials where the behavioral response was correct; young adults showed consistent decreases in N400 component amplitude with additional trials, while older adults did not show this same pattern. Thus, older and younger normal-hearing adults showed similar patterns of improvement on the behavioral measure of word identification over trials. The younger adults also showed evidence that, for the trials that were correct, the cortical resources necessary to generate a correct response were smaller over time. The older adults did not show these cortical-level changes, but they did show behavioral signs of reduced effort associated with their correct responses over time.

Overall, these findings indicate that rapid adaptation to speech produced by NNE talkers is intact in older adults with normal hearing, with younger and older listeners able to achieve a similar rate and magnitude of adaptation, as measured by sentence and word recognition. Both NH listener groups also showed similar patterns of generalization to unfamiliar talkers and accents. The only age-related differences that were seen in this project emerged in Experiment 3, and suggest that aging affects the various components of the non-native speech recognition and adaptation processes – acoustic feature extraction, lexical access, listening effort – differently.

Effects of age-related hearing loss

Only one of the experiments included in this project examined the combinatory effects of aging and age-related declines in sensory acuity on recognition of and adaptation to NNE speech. In Experiment 1, older adults with and without hearing impairment completed a series of five conditions, in which the patterns of rapid adaptation to speech presented in background noise were examined. When the two older listener groups were compared, the main differences were seen in the overall levels of speech recognition performance: the older adults with hearing impairment displayed overall higher speech recognition performance in all but one condition. When the time courses of adaptation were compared across the two older listener groups, no significant differences in curve patterns were observed. Similarly, there were no notable effects of hearing loss on the magnitude of adaptation or generalization to unfamiliar talkers. There were differences between the two older listener groups in terms of the effect of inhibitory mechanisms on rapid adaptation, however the source of these differences is unclear. Older adults with and without hearing impairment both showed similar magnitudes of improvement on nearly all listening conditions, and showed similar patterns of generalization to unfamiliar talkers with familiar and unfamiliar native accents. Together, these results suggest that rapid adaptation ability was not hindered in older adults with hearing impairment. This is a promising finding for the development of auditory training strategies, which may aim to take advantage of rapid adaptation for facilitating learning for challenging speech.

The primary difference seen between the two older listener groups was the overall level of performance. This likely relates to the differences in access to acoustic detail between the two listener groups. As will be discussed in the following section, varying the acoustic features of the stimuli had an effect on the overall levels of performance, but did not seem to have an effect on the patterns of adaptation. While the observed group-wise effect appeared to occur in reverse (OHI listeners often outperformed ONH listeners), this is likely explained by methodological details. The two older listener groups completed the tasks at different signal-to-noise ratios, a methodological choice designed to equate the groups for starting performance in each talker condition. Thus, while it would theoretically be expected that OHI listeners would have poorer performance due to reduced sensory acuity, the ONH listeners may have unintentionally been disadvantaged by the more difficult SNR used during testing, leading to poorer overall performance but similar adaptation patterns.

Effects of individual cognitive capacity

Across the three studies included in this project, several cognitive domains were assessed for their contributions to individual differences in recognition of and adaptation to non-native speech. Broadly, the cognitive domains that were tested include working memory, inhibitory control, attention, executive function, and vocabulary size. These individual characteristics have previously been associated with recognition of challenging speech (Akeroyd, 2008; Bent et al., 2016; Füllgrabe et al., 2015; McLaughlin et al., 2018), as well as aspects of rapid adaptation (Banks et al., 2015; Colby et al., 2018).

Collectively, the present studies failed to identify a cognitive domain that was consistently associated with recognition or rapid adaptation to non-native speech. In Experiment 1, individual differences in inhibitory mechanisms, as measured via the Stroop task, were found to predict recognition of non-native speech. Poorer performance on the Stroop task was associated with poorer speech recognition scores. Stroop scores also interacted with age group and experimental condition to predict patterns of adaptation This is consistent with prior studies documenting the influence of inhibitory control on speech recognition (Dey & Sommers, 2015; Janse, 2012; Sommers & Danielson, 1999) and adaptation (Banks et al., 2015). However, inhibitory control was not found to be a significant predictor of performance in Experiments 2 and 3. These inconsistent findings may relate to differences in procedures across experiments. For example, in Experiment 2, listeners typed in their responses, rather than repeating sentences aloud. Perhaps the processing demands associated with repetition are such that the reliance on inhibitory mechanisms is greater than with a written transcription task. In Experiment 3, inhibitory control was tested using the Flanker task rather than the Stroop task. The Flanker task involves a response to images of arrows, whereas the Stroop task involves a response to written words. It may be that the inherent lexical features in the Stroop task are more likely to involve similar processing resources to those required for a speech recognition task, unlike the Flanker task, which is devoid of lexical information.

Working memory was found to be a significant predictor of event-related potential amplitudes in Experiment 3, and interacted with talker type. Working memory has been associated with N400 effect amplitudes in prior research (Federmeier & Kutas, 2005; Van Petten et al., 1997). Working memory was also expected to be predictive of recognition and adaptation in Experiments 1 and 3, but this was not found to be the case. One potential explanation for this discrepancy could be the differences in stimuli between Experiments 1 and 3. In Experiments 1 and 2, listeners were presented with sentences and were scored on keyword correct performance, whereas in Experiment 3, listeners were instructed to identify only the final target word of each sentence. This difference may have caused a more explicit need for the listeners to store and manipulate the carrier sentence in order to identify the target word, thus possibly relying more heavily on working memory capacity in this Experiment than in Experiment 1.

The non-significant relationships of the other cognitive domains with recognition and adaptation were unexpected. For example, vocabulary size did not emerge as a significant predictor in any of the Experiments, despite the significant relationship documented in prior research (Bent et al., 2016;

Colby et al., 2018; Janse and Adank, 2012) and the importance of lexical information in supporting recognition and adaptation. However, the lack of significant findings in the current experiments cannot be interpreted to mean that receptive vocabulary size does not contribute to individual differences in speech recognition or adaptation performance for foreign-accented speech, but rather that the relationship was not observed with the specific vocabulary measure and the specific listening conditions tested in this study.

Talker and stimulus variability

Throughout this project, a variety of talker conditions were assessed for the listener's ability to recognize and adapt to non-native speech. These conditions allowed for a series of comparisons between conditions varying in type and number of talkers. Variability in the speech stimulus is known to impact speech recognition performance, with recognition typically lower for multiple-talker vs single-talker conditions (Mullennix et al., 1989; Sommers et al., 1994). Some studies have found that variability within the target speech stream results in a slower rate or reduced magnitude of rapid adaptation (Luthra et al., 2021; Wade et al., 2007), though others have not found this to be the case (Tzeng et al., 2016; Witteman et al., 2014). However, variability within the target stimulus has also been shown to be beneficial in facilitating generalization of learning to untrained stimuli (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Lively et al., 1993). One goal of the present set of experiments was to further examine the effect of stimulus variability on rate and magnitude of rapid adaptation as well as generalization, and to assess whether any stimulus effects interacted with aging and/or age-related hearing loss.

Single talker comparisons: Single ENG vs Single SPA

All three experiments included a comparison of performance on simple, semantically meaningful sentences with a single native English talker (ENG) and a single talker producing Spanish-accented English (SPA). The three experiments produced slightly different results when considering overall performance levels, rate and magnitude of adaptation, and generalization.

In Experiments 1 and 2, the presence of a non-native accent in the target signal resulted in different overall performance levels, but had no effect on the rate of adaptation. Similarly, there were no differences in rate of adaptation between the two talkers in Experiment 3. Overall performance levels differed, but only in low-predictability conditions. Magnitude of adaptation was similar for the two single talkers in Experiment 1, while in Experiment 2, the magnitude of adaptation was larger for the SPA than the ENG talker.

Comparisons of single vs multiple talkers and accents

Experiments 1 and 2 additionally contained conditions with single vs multiple non-native talkers producing accented English speech. It was hypothesized that the conditions using multiple talkers would lead to performance declines, particularly if those talkers did not share a language background. This hypothesis was based on a literature documenting poorer speech recognition performance for stimulus lists with multiple interleaved talkers, as compared to single talkers (Goldinger et al., 1991; Mullennix et al., 1989; Pisoni, 1997), including a handful of studies examining this effect with non-native talkers (Bent & Frush Holt, 2013; Kapolowicz et al., 2018). However, it was also hypothesized that listening to multiple talkers would benefit listeners in terms of generalization of learning. Generalization is often more robust following a period of exposure to multiple talkers and/or more variable training stimuli (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Tzeng et al., 2016).

In Experiment 1, two such comparisons were made. First, listeners' performance was compared for a single NS talker (C2) and multiple NS talkers (C4). Next, listeners' performance was compared between multiple NS talkers (C4) and multiple talkers from different language backgrounds (C5). For both comparisons, significant effects of talker were seen when considering overall speech recognition performance for most conditions and groups (though interestingly, this effect occurred in the opposite of the hypothesized direction for the C4/C5 comparison). The exception was the OHI listener group, who did not show an overall talker effect when comparing the single and multiple talker NS talker conditions. However, patterns of adaptation were similar regardless of talker variability for all listener groups. In Experiment 2, two NNE conditions were compared: Single NS and Multiple L1. In this comparison, there were no significant differences in either overall performance level or patterns of adaptation.

The lack of significant differences in patterns of adaptation between single and multiple talker/accent conditions was unexpected. Prior work has shown that an increase in number of talkers leads to a slowed rate of adaptation to nonnative speech (Luthra et al., 2021; Wade et al., 2007). A higher degree of acoustic variability present in the stimulus can slow a listener's rate of learning by requiring the listeners to gain a relatively greater degree of flexibility in their internal category boundaries and stored lexical representations. As Luthra et al. (2021) point out, the addition of multiple talkers requires listeners to continuously update multiple internal models throughout the course of rapid adaptation, which can have greater processing costs as compared to updating a single model. This slowed rate of learning was not seen in the present study. One possible explanation for the contrast with the prior studies is that both Wade et al. (2007) and Luthra et al. (2021) examined learning for phoneme-level stimuli, whereas the present experiments utilized sentence-length materials. As noted below, the presence of lexical information within the stimuli supports rapid adaptation (e.g. Davis et al., 2005); perhaps the detriment of stimulus variability can be overcome by lexical support in sentence-length stimuli as compared to phonemes or words in isolation.

Generalization of learning was seen in both Experiments 1 and 2, when listeners were tested with unfamiliar talkers who shared a language background with a talker that they had heard during the adaptation phase. In Experiment 1, generalization was stronger following adaptation to multiple talkers, as compared to the single-talker conditions. This pattern was not seen in Experiment 2, where similar patterns of generalization were seen in both a single and a multiple-talker condition. Taken together, these findings indicate that listeners can generalize their learning to an unfamiliar talker with a familiar accent, following a period of rapid adaptation. From these results, it is unclear whether the benefit of listening to multiple non-native talkers in facilitating generalization is due to an increased flexibility of internal category boundaries (as suggested by Baese-Berk et al., 2013), or due to a greater likelihood of acoustic similarity between adaptation and generalization talkers (as indicated by Reinisch and Holt, 2013; Xie and Myers, 2017). However, there was no evidence of generalization to an unfamiliar talker with an unfamiliar accent, as tested in Experiment 1, suggesting that even if exposure to multiple talkers did result in broadened internal boundaries, this retuning likely occurred in an accent-specific manner.

Semantic context

The level of semantic context available in the speech stimulus was varied in Experiments 2 and 3. In Experiment 2, listeners were presented with sentences that were either semantically anomalous, or were standard and semantically meaningful. When these two context levels were compared, there was a significant change in the time-course pattern of rapid adaptation to the stimulus. Rapid adaptation to a list of standard, semantically plausible sentences was significantly faster than with semantically anomalous sentences. This was true regardless of the stimulus type. This finding supports prior literature documenting the benefit of lexical information in guiding auditory perceptual learning (Cooper & Bradlow, 2016; Davis et al., 2005; Hervais-Adelman et al., 2011; Jesse & McQueen, 2011; Maye et al., 2008; Norris et al., 2003). Models of perceptual learning for acoustically ambiguous or challenging speech (Norris et al., 2003) indicate that lexical information influences changes to lower-level processing over the course of learning. The findings of Experiment 2 suggest that, in conditions containing lexically intact items (i.e. real words), semantic congruity increases the rate of learning for non-native speech. Perhaps introduction of semantic incongruity incurs a processing cost that disrupts but does not eliminate lexically guided learning.

This context-based difference in the pattern of adaptation was not seen in Experiment 3. In this experiment, listeners were asked to identify target words that either had high or low predictability based on their carrier sentences. Listeners improved their recognition at the same rate regardless of context level or talker accent, though context level and talker did interact in predicting overall performance levels. However, the contextual manipulation and methodologies differ between Experiments 2 and 3.

In Experiment 2, the ‘low’ context condition included semantically anomalous sentences, and listeners were scored on all keywords from the sentence. In Experiment 3, the low-predictability sentence frames were semantically intact, but contained no cues as to the target word of the sentence. Thus, the comparisons indicating a ‘context benefit’ for these two conditions describe somewhat distinct differences. One possible explanation for this distinction may be the fact that both the HP and LP sentence frames utilized in Experiment 3 contained semantically plausible, lexically intact information. Luthra et al. (2020) made a similar comparison of highly predictable and neutral sentence frames in examining lexical retuning of an ambiguous phoneme, and found that both sentence frames provided a similar degree of benefit as compared to no sentence frame at all in facilitating learning.

Interactions of acoustic and semantic features

It was expected that the benefit of context in promoting faster perceptual learning would vary depending on the stimuli. The prior literature that addresses these questions indicates that the benefit of contextual information is dependent on the acoustic features of the target stimulus. For example, Aydelott et al. (2006) found that a context benefit present under intact acoustics was reduced when the stimuli were low-pass filtered. Similarly, Goy et al. (2013) found that the benefit of context was reduced for sentences that had been low-pass filtered, presented in multi-talker babble, or time-compressed. In the present set of studies, no interaction between stimulus type and context level was seen in predicting the patterns of rapid adaptation to non-native English speech.

In comparing the current findings to the prior literature, an important distinction should be made between a speech signal that has been *distorted* versus one that has been acoustically *altered*. The stimuli utilized by Aydelott et al. (2006) and Goy et al. (2013) involve acoustic distortions to the signal: low-pass filtering reduces the spectral information contained in the signal, time-compression reduces the temporal fidelity of the signal, etc. These signal distortions are artificial manipulations employed by researchers in order to answer specific questions about the contributions of temporal or spectral information to speech recognition. In contrast, non-native accent is a naturalistic alteration to the acoustic signal that occurs as a result of the interactions between the talker's native language and the language in which they are speaking. While this interaction results in some alterations in production as compared to a native speaker, it does not result in a loss of temporal or spectral information.

Listeners experience non-native accents in real world listening environments. Recruitment for the present experiments was designed to only include participants who had minimal experience with non-native speech. However, it is extremely rare to find a listener who has never heard a foreign accent, let alone a regional accent or dialect, particularly for in-person experiments that often draw from the University community. Listeners may be more equipped to listen and adapt to challenging speech that occurs in a naturally occurring form as compared to a laboratory-manipulated form, and thus are less affected in their ability to make use of semantic information in these naturalistic listening conditions.

Conclusion

Taken together, the results of this series of experiments indicate that both stimulus variability and semantic context influence recognition of and rapid adaptation to non-native English sentences. These two factors impact different aspects of the recognition and adaptation processes, and appear to act independently. Changes to the level of variability present in the stimulus affected overall levels of performance, but increased stimulus variability did not slow the rate of rapid adaptation to non-native speech. Semantic incongruity slowed the rate of adaptation, but did not reduce the magnitude of learning nor the ability to generalize to unfamiliar talkers. Level of semantic context did not influence rates of rapid adaptation in conditions containing semantically plausible sentences that varied in the degree of supportive semantic context. Younger and older adults with normal hearing showed similar rates of rapid adaptation when measured behaviorally, though older adults did not show the same electrophysiological evidence of greater ease of lexical access over time that was seen in younger adults. Collectively, these findings show that rapid adaptation is intact in younger and older adults even under challenging acoustic conditions. These findings suggest that older adults who report difficulty understanding non-native English speech may be able to benefit from targeted training programs designed to elicit both short-term learning and transfer of learning to untrained non-native talkers, utilizing stimuli containing rich semantic contextual information.

Bibliography

Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*(3), 736–740. https://doi.org/10.1037/a0020054

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299. https://doi.org/10.1098/rstb.2008.0253

Ahissar, Merav, & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

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*(sup2), S53–S71. https://doi.org/10.1080/14992020802301142

Alexander, J. E. D., & Nygaard, L. C. (2019). Specificity and generalization in perceptual adaptation to systematic variation in speech. *The Journal of the Acoustical Society of America*, *145*(3), 1689–1689. https://doi.org/10.1121/1.5101192

Amitay, S., Hawkey, D. J. C., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. In *Perception and Psychophysics* (Vol. 67, Issue 4, pp. 691–698). Psychonomic

Society Inc. https://doi.org/10.3758/BF03193525

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., & Kraus, N. (2013). Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America*, *133*(5), 3030–3038. https://doi.org/10.1121/1.4799804

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, *300*, 18–32. https://doi.org/10.1016/j.heares.2013.03.006

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

Atagi, E., & Bent, T. (2011). Perceptual dimensions of nonnative speech. *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, *August*, 260–263.

Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics*, *41*(6), 509–519. https://doi.org/10.1016/j.wocn.2013.09.003

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory*, *9*(3), 138–150. https://doi.org/10.1101/lm.46502

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29–56. https://doi.org/10.1080/01690960344000099

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*(5), 454–464. https://doi.org/10.1111/j.1469-8986.2006.00448.x

Babel, M., McAuliffe, M., Norton, C., Senior, B., & Vaughn, C. (2019). The goldilocks zone of perceptual learning. *Phonetica*, *76*(2–3), 179–200. https://doi.org/10.1159/000494929

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

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

Behrman, A., & Akhund, A. (2013). The Influence of Semantic Context on the Perception of Spanish-Accented American English. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1567–1578. https://doi.org/10.1044/1092-4388(2013/12-0192)

Bench, J., Kowal, A., & Bamford, J. (1979). The bkb (bamford-kowal-bench) sentence lists for partially-hearing children'. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context. *Ear and Hearing*, *33*(2), 250–256. https://doi.org/10.1097/aud.0b013e31822f680f

Bent, T., & Frush Holt, R. (2013). The influence of talker and foreign-accent variability on spoken word identification. *The Journal of the Acoustical Society of America*, *133*(3), 1677–1686. https://doi.org/10.1121/1.4776212

Bent, T., Holt, R. F., Miller, K., & Libersky, E. (2019). Sentence context facilitation for children's and adults' recognition of native-and nonnative-accented speech. *Journal of Speech, Language, and Hearing Research*, *62*(2), 423–433. https://doi.org/10.1044/2018_JSLHR-H-18-0273

Bieber, R. E., & Gordon-Salant, S. (2017). Adaptation to novel foreign-accented speech and retention of benefit following training: Influence of aging and hearing loss. *Journal of the Acoustical Society of America*, *141*(4), 2800–2811. https://doi.org/10.1121/1.4980063

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, *27*(1), 32–48. https://doi.org/10.1044/jshr.2701.32

Borrie, S. A., Lansford, K. L., & Barrett, T. S. (2017). Generalized adaptation to dysarthric speech. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3110–3117. https://doi.org/10.1044/2017_JSLHR-S-

17-0127

Bradlow, A. (n.d.). *OSCAAR - The Online Speech/Corpora Archive & Analysis Resource*. Retrieved March 26, 2020, from https://oscaar3.ling.northwestern.edu/#!/about

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify english /r/and /1/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

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

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. (2003). Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills*, *97*(1), 11–20. https://doi.org/10.2466/pms.2003.97.1.11

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2017.08.003

Cassarly, C., Matthews, L. J., Simpson, A. N., & Dubno, J. R. (2020). The Revised Hearing Handicap Inventory and Screening Tool Based on Psychometric Reevaluation of the Hearing Handicap Inventories for the Elderly and Adults. *Ear & Hearing*, *41*(1), 95–105.

https://doi.org/10.1097/AUD.0000000000000746

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

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(3), 207–239. https://doi.org/10.1177/00238309040470030101

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

Cooper, A., & Bradlow, A. R. (2016). Linguistically guided adaptation to foreign-accented speech. *The Journal of the Acoustical Society of America*, *140*(5), EL378–EL384. https://doi.org/10.1121/1.4966585

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/2005-04168-006 [pii]\n10.1037/0096-3445.134.2.222

Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychology and Aging*, *30*(3), 634–646. https://doi.org/10.1037/pag0000033

Dimitrijevic, A., Pratt, H., & Starr, A. (2013). *Auditory cortical activity in normal hearing subjects to consonant vowels presented in quiet and in noise*. https://doi.org/10.1016/j.clinph.2012.11.014

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*(1), 538–546. https://doi.org/10.1121/1.428322

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). *Effects of age and mild hearing loss on speech recognition in noise*. *76*(July 1984), 87–96.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914–927. https://doi.org/10.1037/0096-1523.23.3.914

Eckert, M. A., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2008). Age-related effects on word recognition: Reliance on cognitive control systems with structural declines in speech-responsive cortex. *JARO - Journal of the Association for Research in Otolaryngology*, *9*(2), 252–259. https://doi.org/10.1007/s10162-008-0113-3

Eisner, F., & McQueen, J. M. (2005). Specificity of perceptual learning. *Perception & Psycophysics*, *67*(2), 244–238.

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, *7*(DEC), 116. https://doi.org/10.3389/fnsys.2013.00116

Federmeier, K. D., & Kutas, M. (1999). *A rose by any other name: Long-term memory structure and sentence processing*.

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141. https://doi.org/10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., McLennan, D. B., de Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1017/S0048577202001373

Federmeier, K. D., Schwartz, T. J., Van Petten, C., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*, *18*(4), 858–872. https://doi.org/10.1037/0882-7974.18.4.858

Ferguson, M. A., & Henshaw, H. (2015). Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Frontiers in Psychology*, *6*(JAN), 556.

https://doi.org/10.3389/fpsyg.2015.00556

Ferguson, M. A., Henshaw, H., Clark, D. P. A., & Moore, D. R. (2014). Benefits of phoneme discrimination training in a randomized controlled trial of 50- to 74-year-olds with mild hearing loss. *Ear and Hearing*, *35*(4), e110. https://doi.org/10.1097/AUD.0000000000000020

Ferguson, S. H., Jongman, A., Sereno, J. A., & Keum, K. A. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *Journal of the American Academy of Audiology*, *21*(3), 153–162. https://doi.org/10.3766/jaaa.21.3.3

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*(1), 70–79. https://doi.org/10.1121/1.396876

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*, *7*(JAN), 347. https://doi.org/10.3389/fnagi.2014.00347

Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology*, *07*(AUG), 1268. https://doi.org/10.3389/fpsyg.2016.01268

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and*

Performance, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*(3), 325–349. https://doi.org/10.1006/jmla.2000.2741

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the Nature of Talker Variability Effects on Recall of Spoken Word Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152–162. https://doi.org/10.1037/0278-7393.17.1.152

Goldstone, R. (1998). Perceptual learning. *Annual Reviews in Psychology*, *49*, 585–612. https://doi.org/0066-4308/98/0201-0585$08.00 585

Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America*, *121*(3), 1701–1708. https://doi.org/10.1121/1.2436635

Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, *115*(4), 1808–1817. https://doi.org/10.1121/1.1645249

Gordon-Salant, S., & Friedman, S. A. (2011). Recognition of rapid speech by blind and sighted older adults. *Journal of Speech, Language, and Hearing Research*, *54*, 622–631. https://doi.org/10.1044/1092-4388(2010/10-0052)

Gordon-Salant, S., Yeni-Komshian, G. H., & Fitzgibbons, P. J. (2010a). Recognition of accented English in quiet by younger normal-hearing listeners and older listeners with normal-hearing and hearing loss. *The Journal of the Acoustical Society of America*, *128*(1), 444–455. https://doi.org/10.1121/1.3397409

Gordon-Salant, S., Yeni-Komshian, G. H., & Fitzgibbons, P. J. (2010b). Recognition of accented English in quiet and noise by younger and older listeners. *The Journal of the Acoustical Society of America*, *128*(5), 3152–3160. https://doi.org/10.1121/1.3495940

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Schurman, J. (2010). Short-term adaptation to accented English by younger and older adults. *The Journal of the Acoustical Society of America*, *128*(4). https://doi.org/10.1121/1.3486199

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Gosselin, L., Martin, C. D., Navarra-Barindelli, E., & Caffarra, S. (2021). The presence of a foreign accent introduces lexical integration difficulties during late semantic processing. *Language, Cognition and Neuroscience*.

https://doi.org/10.1080/23273798.2021.1909084

Goy, H., Pelletier, M., Coletta, M., & Pichora-Fuller, M. K. (2013). The effects of semantic context and the type and amount of acoustic distortion on lexical decision by younger and older adults. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1715–1732. https://doi.org/10.1044/1092-4388(2013/12-0053)

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205–228. https://doi.org/10.1017/S0142716400002034

Hanulíková, A., van Alphen, P. M., Goch, M. M., & Weber, A. (2012). When one person’s mistake is another’s standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, and Psychophysics*, *74*(3), 613–629. https://doi.org/10.3758/s13414-011-0259-7

Harel-Arbeli, T., Wingfield, A., Palgi, Y., & Ben-David, B. M. (2021). Age-

Related Differences in the Online Processing of Spoken Semantic Context and the Effect of Semantic Competition: Evidence From Eye Gaze. *Journal of Speech, Language, and Hearing Research*, 1–13. https://doi.org/10.1044/2020_JSLHR-20-00142

Hartman, M., & Hasher, L. (1991). Aging and suppression: memory for previously relevant information. *Psychology and Aging*, *6*(4), 587–594.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*(398), 371–386. https://doi.org/10.1080/01621459.1987.10478440

Heinrich, A., & Knight, S. (2016). The contribution of auditory and cognitive factors to intelligibility of words and sentences in noise. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing* (pp. 37–45). Springer.

Henshaw, H., & Ferguson, M. (2014, January 1). Assessing the benefits of auditory training to real-world listening: identifying appropriate and sensitive outcomes. *Proceedings of ISAAR 2013*.

Hervais-Adelman, A., Davis, M. H., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460–474. https://doi.org/10.1037/0096-1523.34.2.460

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., &

Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 283–295. https://doi.org/10.1037/a0020772

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). Routledge.

Huyck, J. J., Smith, R. H., Hawkins, S., & Johnsrude, I. S. (2017). Generalization of perceptual learning of degraded speech across talkers. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3334–3342. https://doi.org/10.1044/2017_JSLHR-H-16-0300

Ingvalson, E. M., Lansford, K. L., Fedorova, V., & Fernandez, G. (2017). Cognitive factors as predictors of accented speech perception for younger and older adults. *The Journal of the Acoustical Society of America*, *141*(6), 4652–4659. https://doi.org/10.1121/1.4986930

Institute of Electrical and Electronics Engineers. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

Irvine, D. R. F., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *The Journal of the Acoustical Society of America*, *108*(6), 2964–2968. https://doi.org/10.1121/1.1323465

Janse, E. (2012a). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology, and*

Cognition, *19*(6), 741–758. https://doi.org/10.1080/13825585.2011.652590

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

Janse, E., & Adank, P. (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

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

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin and Review*, *18*(5), 943–950. https://doi.org/10.3758/s13423-011-0129-2

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. In *Neuropsychology Review* (Vol. 17, Issue 3, pp. 213–233). Springer. https://doi.org/10.1007/s11065-007-9040-z

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*(5), 1337–1351. https://doi.org/10.1121/1.381436

Kapolowicz, M. R., Montazeri, V., & Assmann, P. F. (2018). Perceiving foreign-accented speech with decreased spectral resolution in single- and multiple-talker conditions. *The Journal of the Acoustical Society of America*, *143*(2), EL99–EL104. https://doi.org/10.1121/1.5023594

Karawani, H., Bitan, T., Attias, J., & Banai, K. (2016). Auditory perceptual learning in adults with and without age-related hearing loss. *Frontiers in Psychology*, *6*, 2066. https://doi.org/10.3389/fpsyg.2015.02066

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*(2), 262–268. https://doi.org/10.3758/BF03193841

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology*, *108*, 456–471.

Lacombe, J., Jolicoeur, P., Grimault, S., Pineault, J., & Joubert, S. (2015). Neural changes associated with semantic processing in healthy aging despite intact behavioral performance. *Brain and Language*, *149*, 118–

127. https://doi.org/10.1016/j.bandl.2015.07.003

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain Lang*, *111*(3), 161–172. https://doi.org/10.1016/j.bandl.2009.08.007

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. In *Nature Reviews Neuroscience* (Vol. 9, Issue 12, pp. 920–933). Nature Publishing Group. https://doi.org/10.1038/nrn2532

Lev-Ari, S. (2014). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*(OCT), 1–12. https://doi.org/10.3389/fpsyg.2014.01546

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism*, 1–7. https://doi.org/10.1017/S1366728918001153

Lipnicki, D. M., Crawford, J. D., Dutta, R., Thalamuthu, A., Kochan, N. A., Andrews, G., Lima-Costa, M. F., Castro-Costa, E., Brayne, C., Matthews, F. E., Stephan, B. C. M., Lipton, R. B., Katz, M. J., Ritchie, K., Scali, J., Ancelin, M. L., Scarmeas, N., Yannakoulia, M., Dardiotis, E., … Sachdev, P. S. (2017). Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study. *PLoS Medicine*, *14*(3). https://doi.org/10.1371/journal.pmed.1002261

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Yoh'ichi, T., & Yamada, T. (1994). Training Japanese listeners to identify english /r/ and /l/. iii. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luthra, S., Magnuson, J. S., & Myers, E. B. (2020). Boosting lexical support does not enhance lexically guided perceptual learning. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000945

Luthra, S., Mechtenberg, H., & Myers, E. B. (2021). Perceptual learning of multiple talkers requires additional exposure. *Attention, Perception, and Psychophysics*, 1–12. https://doi.org/10.3758/s13414-021-02261-w

Manheim, M., Lavie, L., & Banai, K. (2018). Age, hearing, and the perceptual learning of rapid speech. *Trends in Hearing*, *22*, 1–18.

https://doi.org/10.1177/2331216518778651

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mattys, S. L., & Scharenborg, O. (2014). Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors. *Psychology and Aging*, *29*(1), 150–162. https://doi.org/10.1037/a0035387

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357

McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America*, *140*(3), 1727–1738. https://doi.org/10.1121/1.4962529

Mcgarr, N. S. (1983). The intelligibility of deaf speech to experience and inexperienced listeners. *Journal of Speech and Hearing Research*, *26*, 451–458. https://doi.org/10.1044/jshr.2603.451

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, *11*(4), 817–822. https://doi.org/10.1097/00001756-200003200-00032

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329–335. https://doi.org/10.1037/h0062491

Mullennix, J. W., Pisoni, D. B., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*(1), 365–378. https://doi.org/10.1121/1.397688

Munro, M. J., & Derwing, T. M. (1995). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Nasreddine, Z. S., Phillips, N. A., BÃ©dirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Neger, T. M., Rietveld, T., & Janse, E. (2014). *Relationship between perceptual learning in speech and statistical learning in younger and older adults*. https://doi.org/10.3389/fnhum.2014.00628

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical*

Society of America, *87*(6), 2705–2715. https://doi.org/10.1121/1.399061

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

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*(7), 989–1001. https://doi.org/10.3758/BF03205458

Ortiz, J. A., & Wright, B. A. (2010). Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill. *Experimental Brain Research*, *201*(3), 441–451. https://doi.org/10.1007/s00221-009-2053-5

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. https://doi.org/10.1037/0278-7393.19.2.309

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, *1687*, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

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 ISSNOnline) Journal International Journal of Audiology*, *47*, 1499–2027. https://doi.org/10.1080/14992020802307404

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M., Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Pisoni, D. (1997). Some thoughts on “normalization” in speech perception. In J. Mullennix & K. Johnson (Eds.), *Talker variability in speech processing* (pp. 9–32). Academic Press.

Porretta, V., & Kyröläinen, A. J. (2019). Influencing the Time and Space of Lexical Competition: The Effect of Gradient Foreign Accentedness. *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*(10). https://doi.org/10.1037/xlm0000674

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21. https://doi.org/10.1016/j.wocn.2016.05.006

Reinisch, E., & Holt, L. L. (2013). *Lexically guided phonetic retuning of foreign-accented speech and its generalization*. https://doi.org/10.1037/a0034409

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, *17*(3), 781–791. https://doi.org/10.1016/S0926-6410(03)00202-7

Reitan, R. M. (1971). Trail making test results for normal and brain-damaged children. *Perceptual and Motor Skills*, *33*(2), 575–581. https://doi.org/10.2466/pms.1971.33.2.575

Rogers, C. S., & Wingfield, A. (2015). Stimulus-independent semantic bias misdirects word recognition in older adults. *The Journal of the Acoustical Society of America*, *138*(1), EL26–EL30. https://doi.org/10.1121/1.4922363

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). *Processing changes when listening to foreign-accented speech*. *9*. https://doi.org/10.3389/fnhum.2015.00167

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(July), 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(sup2), S99–S105. https://doi.org/10.1080/14992020802301167

Roque, L., Gaskins, C., Gordon-Salant, S., Goupell, M. J., & Andersona, S. (2019). Age effects on neural representation and perception of silence duration cues in speech. *Journal of Speech, Language, and Hearing Research*, *62*(4), 1099–1116. https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0076

Rossi, S., Hartmüller, T., Vignotto, M., & Obrig, H. (2013). Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. *Brain and Language*, *127*(3), 404–414. https://doi.org/10.1016/j.bandl.2013.02.009

Rotman, T., Lavie, L., & Banai, K. (2020). Rapid Perceptual Learning: A Potential Source of Individual Differences in Speech Perception Under Adverse Conditions? *Trends in Hearing*, *24*. https://doi.org/10.1177/2331216520930541

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review*, *10*(1), 101–124.

https://doi.org/10.1016/0273-2297(90)90006-P

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

Samuel, A., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Scharenborg, O., & Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, and Psychophysics*, *75*(3), 525–536. https://doi.org/10.3758/s13414-013-0422-4

Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception, and Psychophysics*, *77*(2), 493–507. https://doi.org/10.3758/s13414-014-0792-2

Schurman, J., Brungart, D., & Gordon-Salant, S. (2014). Effects of masker type, sentence context, and listener age on speech recognition performance in 1-back listening tasks. *The Journal of the Acoustical Society of America*, *136*(6), 3337–3349. https://doi.org/10.1121/1.4901708

Scott, D., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450–466.

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older listeners. *Journal of the Acoustical Society of America*, *123*(1). https://doi.org/10.1121/1.2783762

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316. https://doi.org/10.1121/1.3101452

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America*, *101*(4), 2278–2288. https://doi.org/10.1121/1.418208

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, *14*(3), 458–472. https://doi.org/10.1037/0882-7974.14.3.458

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, *96*(3), 1314–1324. https://doi.org/10.1121/1.411453

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Sow, A., & Friedman, A. (2020). *Big Friendship: How We Keep Each Other Close*. Simon & Schuster.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G. (1994). Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *92*(1), 30–43. https://doi.org/10.1016/0168-5597(94)90005-1

Stevens, F. L., Hurley, R. A., & Taber, K. H. (2011). Anterior cingulate cortex: Unique role in cognition and emotion. *Journal of Neuropsychiatry and Clinical Neurosciences*, *23*(2), 121–125. https://doi.org/10.1176/jnp.23.2.jnp121

Straus, A., Kotz, S., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the n400. *Journal of Cognitive Neuroscience*, *26*(3), 194–198. https://doi.org/10.1162/jocn

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sumner, M., & Samuel, A. G. (2009). *The effect of experience on the perception and representation of dialect variants*. https://doi.org/10.1016/j.jml.2009.01.001

Taler, V., Aaron, G., Steinmetz, L., & Pisoni, D. (2010). Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *Journal of Gerontology*, *65B*(5), 551–560. https://doi.org/10.1093/geronb/gbq039

Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics and Phonetics*, *9*(2), 139–154. https://doi.org/10.3109/02699209508985329

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, *102*(6), 3762–3773. https://doi.org/10.1121/1.420139

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, A. B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90. https://doi.org/10.1097/00003446-200104000-00001

Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332–1343. https://doi.org/10.1016/S1388-2457(03)00114-7

Tzeng, C. Y., Alexander, J. E. D., Sidaras, S. K., & Nygaard, L. C. (2016). *The role of training structure in perceptual learning of accented speech*. 1–29. https://doi.org/10.1037/xhp0000260

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working Memory Capacity Dissociates Lexical and Sentential Context Effects.

Psychological Science, *8*(3), 238–242. https://doi.org/10.1111/j.1467-9280.1997.tb00418.x

Van Rij, J., Vaci, N., Wurm, L., & Feldman, L. B. (2020). Alternative quantitative methods in psycholinguistics: Implications for theory and design. In V. Pirreli, I. Plag, & W. Dressler (Eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon* (pp. 83–126). De Gruyter.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*(2–3), 122–144. https://doi.org/10.1159/000107913

Wayne, R. V., Hamilton, C., Huyck, J. J., & Johnsrude, I. S. (2016). Working memory training and speech in noise comprehension in older adults. *Frontiers in Aging Neuroscience*, *8*(MAR), 49. https://doi.org/10.3389/fnagi.2016.00049

Weber, A., Betta, A. M. Di, & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*(1), 34–51. https://doi.org/10.1016/j.wocn.2014.05.002

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). *Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli*. *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, *7*(3), 175–182.

Wingfield, A., Alexander, A. H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Language and Speech*, *37 (Pt 3)*(3), 221–235. https://doi.org/10.1177/002383099403700301

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516669723

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, and Psychophysics*, *75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin and Review*, *21*(2), 512–519. https://doi.org/10.3758/s13423-013-0519-8

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/10.1037/a0029206

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, *62*(4), 1025–1036. https://doi.org/10.1111/j.1541-0420.2006.00574.x

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Wright, B. A, & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 301–311. https://doi.org/10.1098/rstb.2008.0262

Wright, Beverly A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, *17*(10), 3956–3963. https://doi.org/10.1523/jneurosci.17-10-03956.1997

Wright, Beverly A., & Sabin, A. T. (2007). Perceptual learning: How much daily training is enough? *Experimental Brain Research*, *180*(4), 727–736.

https://doi.org/10.1007/s00221-007-0898-z

Wright, Beverly A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *Journal of Neuroscience*, *30*(35), 11635–11639. https://doi.org/10.1523/JNEUROSCI.1441-10.2010

Xie, X., & Jaeger, T. F. (2020). Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America*, *147*(5), 3322–3347. https://doi.org/10.1121/10.0001141

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, *97*, 30–46. https://doi.org/10.1016/j.jml.2017.07.005

Zaltz, Y., Kishon-Rabin, L., Karni, A., & Ari-Even Roth, D. (2020). Practice Makes Transfer Imperfect. *Ear and Hearing*, 1. https://doi.org/10.1097/aud.0000000000000860

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*.