ABSTRACT

Title of dissertation:	AGE EFFECTS ON PERCEPTUAL ORGANIZATION OF SPEECH IN REALISTIC ENVIRONMENTS
	William Joseph Bologna, Doctor of Philosophy, 2017
Dissertation directed by:	Professor Sandra Gordon-Salant Department of Hearing & Speech Sciences

Communication often occurs in environments where background sounds fluctuate and mask portions of the intended message. Listeners use envelope and periodicity cues to group together audible glimpses of speech and fill in missing information. When the background contains other talkers, listeners also use focused attention to select the appropriate target talker and ignore competing talkers. Whereas older adults are known to experience significantly more difficulty with these challenging tasks than younger adults, the sources of these difficulties remain unclear. In this project, three related experiments explored the effects of aging on several aspects of speech understanding in realistic listening environments. Experiments 1 and 2 determined the extent to which aging affects the benefit of envelope and periodicity cues for recognition of short glimpses of speech, phonemic restoration of missing speech segments, and/or segregation of glimpses with a competing talker. Experiment 3 investigated effects of age on the ability to focus attention on an expected voice in a two-talker environment. Twenty younger adults and 20 older adults with normal hearing participated in all three experiments and also completed a battery of cognitive measures to examine contributions from specific cognitive abilities to speech recognition. Keyword recognition and cognitive data were analyzed with an item-level logistic regression based on a generalized linear mixed model. Results indicated that older adults were poorer than younger adults at glimpsing short segments of speech but were able use envelope and periodicity cues to facilitate phonemic restoration and speech segregation. Whereas older adults performed poorer than younger adults overall, these groups did not differ in their ability to focus attention on an expected voice. Across all three experiments, older adults were poorer than younger adults at recognizing speech from a female talker both in quiet and with a competing talker. Results of cognitive tasks indicated that faster processing speed and better visual-linguistic closure were predictive of better speech understanding. Taken together these results suggest that age-related declines in speech recognition may be partially explained by difficulty grouping short glimpses of speech into a coherent message, which may be particularly difficult for older adults when the talker is female.

AGE EFFECTS ON PERCEPTUAL ORGANIZATION OF SPEECH IN REALISTIC ENVIRONMENTS

by

William Joseph Bologna

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfilment
Of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:

Professor Judy R. Dubno, Co-Chair
Professor Sandra Gordon-Salant, Co-Chair
Professor Monita Chatterjee
Professor Rochelle S. Newman
Professor Jonathan Z. Simon

For Nancy,

who instilled in me a sense of curiosity and the thirst for knowledge.

TABLE OF CONTENTS

LIST OF FIGURES

I. Literature Review

There is a rich history of literature on age-related declines in speech recognition in realistic listening environments. For older adults with hearing loss, speech recognition difficulty likely reflects the combined effects of age and reduced audibility (CHABA, 1988). When audibility is controlled or accounted for by experimental or statistical methods, residual effects of age on speech recognition persist in complex backgrounds, such as in modulated noise (Dubno, Horwitz, & Ahlstrom, 2002; 2003; Eisenberg, Dirks, & Bell, 1995) and with competing speech (Helfer & Freyman, 2008; Rajan & Cainer, 2008). Several explanations of this residual age-related decline in speech recognition have been proposed, including declines in temporal processing (Fitzgibbons & Gordon-Salant, 1996; Gordon-Salant & Fitzgibbons, 1993), and cognitive decline (Humes, Watson, Christensen, Cokely, Halling, & Lee, 1994; Humes, Kidd, & Lentz, 2013). One explanation that has received less attention is age-related declines in "perceptual organization," or the process by which the auditory system interprets acoustic input and creates an internal representation of an auditory scene (Bregman, 1990). The majority of experiments on this topic have included younger adults with normal hearing and described the roles these processes serve in a normal auditory system. As a result, gaps of knowledge remain regarding the extent to which these processes may change or decline with advancing age and explain speech recognition difficulties experienced by older adults. Here, it is proposed that poorer speech recognition in realistic listening environments by older adults may be attributed, in part, to age-related declines in perceptual organization. This hypothesis is supported by existing psychoacoustic and speech perception literature, which shows that older adults are poorer than younger adults at using acoustic cues in speech that are known to facilitate perceptual organization in complex backgrounds. Age-related changes in auditory function, as well as cognitive changes associated with increasing age, may have implications for the efficiency with which the auditory system can organize incoming acoustic information from multiple sources, thereby limiting speech recognition abilities of older adults in realistic environments.

I.A. Introduction to Perceptual Organization

Realistic listening environments contain sounds from multiple sources and these sounds are mixed together as they enter the ear. Perceptual organization refers to the process of disentangling these sound mixtures and prioritizing the processing of relevant signals in favor of irrelevant signals. Perceptual organization has been described as consisting of two component processes: object formation, which refers to generating separate representations of individual sound sources, and object selection, which refers to choosing a particular sound source as the focus of attention and higher level processing (Shinn-Cunningham, 2008). Object formation is typically described as a lower level auditory process driven by acoustic cues that function similarly to Gestalt grouping rules in vision (Bregman, 1990; Darwin, 1997), whereas object selection is considered a higher-level process driven by the listener’s expectations and intentions (Best, Ozmeral, Kopčo, & Shinn-Cunningham, 2008; Ding & Simon, 2012; Maddox & Shinn-Cunningham, 2012). Object formation and object selection are related, but serve separate and equally important roles in processing information in realistic listening environments. As such, it is important to study object formation and selection separately to fully understand factors that may influence communication in these environments and explain age-related declines (Ihlefeld & Shinn-Cunningham, 2008).

I.B. Object Formation

In order to extract information accurately and efficiently in a complex listening environment, the auditory system generates mental representations of individual sound sources. These mental representations are referred to as “auditory objects,” and “object formation” is the process of decomposing a mixture of sounds into distinct auditory objects (Griffiths & Warren, 2004; Shinn-Cunningham, 2008). Object formation can be further broken down into two component processes: (1) “simultaneous segregation,” which refers to separating sounds from different sources when they occur simultaneously or close together in time and (2) “sequential integration,” which is the process by which successive sounds from a single source are linked together across time (Bregman, 1990; Darwin, 1997). In realistic listening environments, these two processes work together to facilitate speech understanding by separating audible segments of speech from the background and assimilating those audible segments over time into a single auditory object. A deficiency in either or both of these processes may contribute to speech recognition difficulty of older adults in complex backgrounds.

Simultaneous segregation is most often studied using tasks requiring detection or recognition of concurrently presented sounds. Concurrent vowel identification tasks have demonstrated age-related declines in identification as well as declines in behavioral and neural sensitivity to segregation cues, such as fundamental frequency (F0; Arehart, Souza, Muralimanohar, & Miller, 2011; Chintanpalli, Ahlstrom, & Dubno, 2014; 2016; Snyder & Alain, 2005; Summers & Leek, 1998; Vongpaisal & Pichora-Fuller, 2007). Using non-speech sounds, Neff and Green (1987) evaluated simultaneous segregation of a target tone from multi-tone maskers. This paradigm was later expanded by Kidd, Mason, Deliwala, Woods, and Colburn (1994) who determined that repeated presentation of a tonal target in randomly varying multi-tone maskers facilitated detection of the target. In a follow-up study, the advantage of repeated presentations was determined to depend not on a simple accumulation of evidence over "multiple looks," but rather a time-dependent buildup of an auditory object, illustrating that simultaneous segregation can be facilitated by sequential integration in a complex background (Kidd, Mason, & Richards, 2003). One recent study using the multi-tone masking paradigm of Kidd et al. (1994) revealed that detection of a tone in a multi-tone masker was poorer for older than younger adults, suggesting that these critical object formation mechanisms may decline with increasing age (Humes et al., 2013).

Sequential integration can be studied in the absence of simultaneous segregation using repetitive sequences of non-overlapping sounds. When two sounds are presented as a rapid sequence of alternating triplets (ABA-ABA), listeners will perceive two component sequences (A-A-A-A and –B—B-; Bregman, 1990; van Noorden, 1975). Numerous studies of this “streaming effect” have demonstrated that any salient characteristic (e.g., pitch, timbre, loudness, spatial location) may be used as a basis for sequential integration (Moore & Gockel, 2012). However, the extent to which streaming is affected by age remains unclear. Previous studies with simple tonal stimuli have shown no differences between younger and older adults (Alain, Ogawa, & Woods, 1996; Snyder & Alain, 2007; Trainor & Trehub, 1989), whereas more recent work with complex stimuli and tasks have shown poorer sequential integration among older adults compared to younger adults (Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2001; Hutka, Alain, Binns, & Bidelman, 2013; Rimmele, Schröger, & Bendixen, 2012). Thus, it appears that older adults retain the basic ability to form an auditory object through sequential integration of sounds over time, but they may be less adept than younger adults at forming a coherent object when additional challenges are imposed by complex auditory signals and backgrounds. These age-related difficulties in realistic listening environments may result from declines in the ability to use temporal cues in speech to facilitate sequential integration

I.B.1. Temporal cues for object formation

Natural speech contains temporal cues across a broad range of fluctuation rates, each of which contributes slightly different information. The framework described by Rosen (1992) distinguishes between three types of speech cues based on their temporal fluctuation rates. The envelope is described as slow amplitude modulations between rates of 2-50 Hz, which codes syllabic and segmental rates and provides information about rhythm and prosody of speech. Periodicity refers to faster fluctuations in amplitude between rates of 50-500 Hz, which code voicing and pitch information. Fluctuation rates above 500 Hz are referred to as temporal fine-structure (TFS), though the role these cues play in speech recognition is still under debate (c.f. Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Swaminathan & Heinz, 2012). Each of these three cues may contribute to object formation in different ways, particularly in realistic environments where background sounds may disrupt cues at certain rates, but leave others relatively intact.

Envelope cues are likely the most important for facilitating speech recognition, particularly in a quiet environment. This has been demonstrated in many studies in which speech is represented by only the temporal envelope via noise vocoding (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). For noise vocoding, the speech signal is first separated into frequency bands and then the temporal envelope within each band is extracted and used to modulate noise carriers for each band. Once the noise bands are combined, the resulting vocoded speech is a highly intelligible representation of the temporal envelope of the original speech, but without any periodicity information (e.g., Drullman, 1995; Shannon et al., 1995; Van Tasell, Soli, Kirby, & Widin, 1987). Whereas continuity of these slow modulations can provide a structured and predictable pattern to assist sequential integration (Grimault, Bacon, & Micheyl, 2002), envelope cues are susceptible to a variety of masking effects from competing signals. Energetic masking can occur when the target and a masker overlap in both time and frequency, rendering portions of the target's envelope inaudible. Modulation masking may occur when the masker contains amplitude fluctuations that disrupt the listener's ability to distinguish target envelope modulations from those of the masker (Kwon & Turner, 2001; Qin & Oxenham, 2003; Stone, Füllgrabe, & Moore, 2012). These masking effects are particularly problematic for speech recognition in a competing talker background, because the modulation rate of the speech envelope and its pattern across frequencies is inherently similar between talkers. As such, listeners may rely on faster rate cues to help facilitate object formation for speech in complex backgrounds (Stickney, Assmann, Chang, & Zeng, 2007).

Periodicity cues carry less information than the envelope on the content of the message, but they can be used to distinguish between speech segments from different talkers. A speaker's fundamental frequency and intonation are coded by the periodic fluctuations in the range of 50-500 Hz, and these characteristics are typically unique to an individual talker's utterance. As such, these cues are particularly important for resolving informational masking effects, which occur when speech from multiple talkers overlaps across time (Brungart, 2001; Ihlefeld & Shinn-Cunningham, 2008). Listeners can exploit differences in fundamental frequency to facilitate simultaneous segregation of target speech from competing talkers and use similarities in pitch and intonation to assist with sequential integration of speech segments separated in time (Brokx & Nooteboom, 1982; Gaudrain, Grimault, Healy, & Béra, 2007). By contrast, vocoded speech is characterized by an absence of periodicity cues and intelligibility of these signals is substantially reduced when the background consists of competing talkers (Stickney, Zeng, Litovsky, & Assmann, 2004). For older adults, declines in periodicity coding in the brainstem may contribute to reduced ability to use these cues to facilitate speech object formation in a complex background (Clinard & Tremblay, 2013). This hypothesis is supported by several behavioral studies indicating that older adults are more susceptible to informational masking effects than younger adults (e.g., Helfer & Freyman, 2008; Rajan & Cainer, 2008).

The importance of TFS cues for speech recognition remains somewhat unclear. Lorenzi and colleagues synthesized speech to retain only TFS cues and demonstrated greatly reduced intelligibility among younger and older adults with hearing loss compared to younger adults with normal hearing (Lorenzi et al., 2006). However, narrowband filtering of TFS speech in the cochlea results in “recovered” envelope cues at the level of the auditory nerve, which may account for the relatively good intelligibility of TFS speech among normal hearing listeners (Ghitza, 2001; Swaminathan & Heinz, 2012). One recent study demonstrated that coherence of TFS information across ears facilitates speech understanding in a competing talker background, suggesting a possible role of TFS for object formation with binaural listening (Swaminathan, Mason, Streeter, Best, Roverud, & Kidd, 2016). Using psychophysical methods, Füllgrabe and colleagues demonstrated poorer TFS sensitivity for older adults compared to younger adults, as well as a correlation between TFS sensitivity and speech recognition

(Füllgrabe, Moore, & Stone, 2014). However, TFS sensitivity was also correlated with cognitive factors that are known to contribute to speech recognition. Whereas some evidence exists to suggest that age-related declines in TFS sensitivity may contribute to object formation difficulty among older adults, it remains largely unclear what role TFS cues play in monaural speech recognition. A comprehensive evaluation of TFS cues is outside the scope of the present work, which will instead focus on the effects of age on the use of envelope and periodicity cues for object formation.

I.B.2. Speech recognition and "glimpsing"

Speech unfolds over time with natural fluctuations and brief silences intrinsic to phonemic and syllabic structure. Each successive segment of speech must be incorporated into a single auditory object. Under optimal listening conditions, continuity of the temporal envelope and periodicity in voiced segments provides redundant cues to facilitate formation of an auditory object (Bregman, 1990). In realistic listening environments, portions of speech are often rendered inaudible by a fluctuating background. Under these conditions, the message must be interpreted based on the remaining audible fragments, or "glimpses" of speech (Buus, 1985; Cooke, 2006; Moore, 2003). To achieve adequate speech understanding in these environments, the auditory system uses the process of sequential integration to incorporate glimpses of speech across gaps of missing information and form a single auditory object (Assmann & Summerfield, 2004; Cooke, 2006). This form of speech-based sequential integration, often referred to as "glimpsing," is likely to be more difficult in the presence of a speech masker than a noise masker, as speech segments from competing talkers may be incorrectly incorporated into the auditory object. Speech recognition by older adults is vulnerable to fluctuating maskers (Dubno et al., 2002; Festen & Plomp, 1990; Takahashi & Bacon, 1992), particularly when those maskers are competing talkers (Duquesnoy, 1983; Rajan & Cainer, 2008; Tun, O'Kane, & Wingfield, 2002). It remains unclear the extent to which difficulty with glimpsing may play a role in age-related declines in speech recognition in fluctuating maskers.

Previous research on recognition of speech in modulated noise provides evidence of an age-related decline in the ability to use glimpses of speech in realistic environments. When masking noise is modulated, either by a square wave or sinusoidal wave, dips in the level of the masker offer opportunities for listeners to glimpse speech at favorable SNRs. As a result, speech recognition in modulated noise is often better than recognition in steady-state noise; the improvement in performance is referred to as “masking release,” and is thought to reflect the benefit associated with glimpsing (see Moore, 1990 for a review). One common finding is that masking release is reduced in older adults, particularly those with peripheral hearing loss (Dubno et al., 2002; Festen & Plomp, 1990; Jin & Nelson, 2006; Takahashi & Bacon, 1992). For listeners with hearing loss, reduced sensation levels of speech glimpses are likely a primary factor limiting masking release (Bacon, Opie, & Montoya, 1998; Festen & Plomp, 1990; Gustafsson & Arlinger, 1994). However, age-related declines in masking release persist even when speech is presented at high sensation levels

(Eisenberg et al., 1995), or when older listeners with normal hearing are compared to younger adults (Dubno et al., 2002). These findings suggest additional factors may limit masking release in older adults, such as reduced temporal resolution (Jin & Nelson, 2006; Takahashi & Bacon, 1992), prolonged recovery from forward masking (Dubno et al., 2003), and declines in sensitivity to TFS cues (Hopkins & Moore, 2009; Lorenzi et al., 2006).

A typical Gaussian noise used in masking release experiments contains random fluctuations in amplitude, which can result in modulation masking of speech in addition to energetic masking (Stone et al., 2012). Modulation masking effects occur because the inherent fluctuations of the noise disrupt the listener's ability to recognize important envelope modulations in the speech signal. When a typical noise masker is modulated to facilitate glimpsing, the momentary reductions in the energy of the masker provide a release from both energetic and modulation masking. Stone and colleagues (2012) processed noise to reduce the inherent envelope fluctuations and found that the resulting "low-noise noise" was a considerably less effective masker of speech and produced minimal masking release when the noise was modulated. These results suggest that modulation masking constitutes a significant portion of the masking effects associated with typical Gaussian noise, and masking release associated with Gaussian noise reflects a release primarily from modulation masking rather than energetic masking.

Studies using the masking release paradigm with vocoded speech provide some evidence on the roles that envelope and periodicity cues may play in glimpsing. Several studies have demonstrated a lack of masking release when vocoded speech is presented in modulated noise (Nelson & Jin, 2004; Nelson, Jin, Carney, & Nelson, 2003; Stickney, et al. 2004). The loss of periodicity cues during the vocoding process, coupled with the use of modulated noise bands to represent the speech envelope, results in perceptual similarity between vocoded speech and a modulated noise masker (Jin, Nie, & Nelson, 2013; Stickney et al., 2004). A consequence of this perceptual similarity may be that listeners are unable to distinguish speech glimpses from the noise segments. As a result, intermittent noise segments may be integrated into the auditory object along with glimpses of speech, thereby disrupting intelligibility of the signal. These results suggest that without periodicity cues to distinguish the target from a masker, envelope cues may not be sufficient for object formation in realistic listening environments. For older adults, declines in sensitivity to periodicity cues may result in a similar difficulty distinguishing target glimpses from the masker, which may compromise object formation in realistic environments where only glimpses of speech are audible.

I.B.3. Interrupted speech and phonemic restoration

Another paradigm for studying sequential integration and glimpsing is to interrupt speech with silence, rather than noise. Jin and Nelson (2006; 2010) compared sentence recognition in modulated noise to recognition of sentences interrupted by silence and found the two measures to be highly correlated. In general, recognition of interrupted speech depends primarily on the proportion of speech that remains after the speech is interrupted (Gilbert, Bergeras, Voillery, &

Lorenzi, 2007; Wang & Humes, 2010). If the proportion of speech is held constant, effects of interruption rate can also be observed for sentence length material (Miller & Licklider, 1950). For example, for a 0.50 proportion of speech remaining, very slow interruption rates (< 1 Hz) will cause whole words to be present during “on” portions, resulting in fragmented sentences in which some words are easily identifiable and others are missing. At much faster rates (> 20 Hz), interruptions are brief enough that “on” portions sample each syllable and listeners can identify syllabic and word level items with relative ease. At moderate rates (between 2-5 Hz), interruptions are frequent enough that whole words are rarely retained and interruptions are long enough in duration that perceived continuity and overall recognition of the sentence declines (Bashford & Warren, 1987; Bashford, Meyers, Brubaker, & Warren, 1988). Older adults typically demonstrate poorer recognition of interrupted speech than younger adults, particularly at these moderate interruption rates where perceived continuity of the sentence is affected (Gordon-Salant & Fitzgibbons, 1993; Krull, Humes, & Kidd, 2013; Shafiro, Sheft, Risley, & Gygi, 2015). Ratings of perceived continuity have been shown to correlate with measures of sequential integration (Bregman, Colantonio, & Ahad, 1999), and improvements in continuity are associated with stronger activations of speech and language areas of the brain and better speech recognition (Heinrich, Carlyon, Davis, & Johnsrude, 2008). These results suggest that poor recognition of interrupted speech among older adults may be a consequence of age-related declines in glimpsing.

In one of the first studies of interrupted speech, Miller and Licklider (1950) reported that when interrupted speech was gated out of phase with interrupted noise (such that the noise was on while the speech was off) listeners experienced a “picket fence” percept, wherein the interrupted sentence was perceived as continuous behind the bursts of noise. The effect is similar to a non-linguistic phenomenon, auditory induction, where a sequence of short tones alternating with noise bursts is perceived as a single continuous tone with an intermittent noise masker (Warren, Obusek, & Ackroff, 1972). In both the speech and non-speech examples, energy from bursts of noise serve as evidence that missing information is being masked, resulting in an illusion of continuity (Warren, 1984). Later explorations of the phenomenon with interrupted speech revealed that for sentence length material, the addition of noise to the silent intervals improves speech recognition; this effect is referred to as “phonemic restoration” (Bashford & Warren, 1987; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983, Warren, 1970). Phonemic restoration and auditory induction share a common interpretation that the illusion of continuity created by the noise bursts allows successive segments to be more easily fused together into a continuous percept. In the case of phonemic restoration, the continuity illusion benefits sequential integration of speech glimpses, which leads to an improvement in sentence recognition.

In contrast to the extensive literature on phonemic restoration in younger adults, relatively few studies have examined the effect in older adults. Başkent, Eiler, and Edwards (2010) examined phonemic restoration in groups of listeners with normal hearing, mild hearing loss, and moderate hearing loss. Though age was not a selection criterion in their study, normal hearing subjects tended to be younger (average age of 37 years) than the subjects with mild and moderate hearing loss (average ages of 70 and 73 years, respectively). Phonemic restoration benefit was similar for subjects with normal hearing and mild hearing loss, but little or no phonemic restoration was observed in subjects with moderate hearing loss. An analysis using the Articulation Index suggested that this finding was independent of differences in audibility and a follow up analysis indicated the lack of phonemic restoration in the moderate hearing loss group was independent of baseline speech recognition scores (Başkent, 2010). However, due to the overlapping age ranges of subjects with mild and moderate hearing loss, these results cannot be definitively linked to aging.

Only one study has investigated age-related changes in phonemic restoration without confounding effects of hearing loss (Saija, Akyurek, Andringa, & Başkent 2014). Saija and colleagues measured phonemic restoration in younger and older adults with normal hearing across a range of interruption rates (0.625-20 Hz). Interrupted segments were either left silent, or filled with steady-state noise at -10 dB SNR (re: “on” portions of speech); phonemic restoration was defined as the difference in recognition scores between silent-interrupted and noise-filled sentences. In addition, sentences were time-compressed or expanded without altering the voice pitch in order to assess phonemic restoration for sentences presented at a slow, normal, or fast rate. Their results indicated that older adults benefited more from phonemic restoration than younger adults for moderate interruption rates (~2.5 Hz) in sentences presented at a normal or slow rate. The authors interpreted this finding as an indication that older listeners with normal hearing are more adept at filling in missing information in speech. However, recognition of silence-interrupted speech is known to be poorer for older adults than younger adults (Gordon-Salant & Fitzgibbons, 1993; Krull et al., 2013; Shafiro et al., 2015), which is the condition that serves as the baseline for measures of phonemic restoration. As such, age-related declines in the ability to connect glimpses of speech across gaps of silence may also contribute to apparent enhanced phonemic restoration in older adults. The addition of noise to silent intervals improves continuity and may facilitate better sequential integration of speech glimpses, relative to silence-interrupted speech. As a result, larger phonemic restoration benefit for older adults may be observed as a consequence of their poorer performance for the baseline condition. In this way, phonemic restoration may reflect a form of perceptual scaffolding, where the addition of noise helps listeners form a coherent auditory object from glimpses of speech separated in time. Older adults may benefit more from this supportive mechanism than younger adults, due to age-related declines in the ability to connect short segments of speech across silent intervals. This hypothesis warrants further investigation and serves as a primary motivation for Experiment 1.

I.B.4. Lexical factors and phonemic restoration

Another explanation for increased phonemic restoration among older adults is that a longer lifetime of exposure to language helps older adults fill in missing information. Older adults are known to benefit more from supportive contextual cues on sentence recognition tasks (Pichora-Fuller, 2008). The presentation of partial linguistic information in sentences may allow older adults to leverage their language abilities in a similar way, leading to greater increases in recognition via phonemic restoration. This may facilitate restoration of more linguistically difficult stimulus items among older adults, compared to younger adults.

Several factors can influence the linguistic difficulty of stimulus words on a speech recognition task (i.e., Luce & Pisoni, 1998). Word frequency is a measure of how commonly a word is used in spoken language; words that occur with greater frequency are typically recognized more easily (Howes, 1954). A word’s neighborhood density refers to the number of similar sounding words that exist in the language, which is typically quantified as words that differ from the stimulus word by only 1 phoneme. Words with more lexical neighbors are more likely to be misheard as similar-sounding words (Cluff & Luce, 1990). Phonotactic probability is the relative frequency of occurrence for sequences of phonemes in the word. A common measure is biphone probability, which is calculated based on the frequencies of phoneme pairs within the stimulus; words with greater biphone probability are typically recognized more easily (Vitevitch, Luce, Pisoni, & Auer, 1999). These lexical characteristics may influence the extent to which missing portions of a word can be perceptually repaired via phonemic restoration.

The contribution of lexical factors on phonemic restoration is largely unknown. Samuel (1981a, b) included a measure of word frequency for stimulus words on a phonemic restoration task, and found only modest effects of word frequency on phonemic restoration. However, phonemic restoration in these studies was not measured via speech recognition. Rather, Samuel measured d’ for distinguishing between noise-replaced vs. noise-masked syllables in stimulus words. An evaluation of lexical factors for recognition of silence-interrupted and noise-filled sentences is warranted to further determine the contributions of lexical factors to phonemic restoration.

I.B.5. Phonemic restoration with envelope and periodicity cues

Results of two studies have demonstrated that modulating intervening noise by the envelope of the missing speech segment enhances the phonemic restoration effect. Bashford, Warren, and Brown (1996) referred to these envelope modulations as bottom-up cues that provide beneficial information to assist the restoration mechanism, whereas Shinn-Cunningham and Wang (2008) interpreted their result as an indication that noise modulations were incorporated into the speech object. Both accounts suggest that continuity of envelope cues across the duration of an interrupted sentence improves the integration of speech glimpses into a coherent and intelligible auditory object. This effect has only been studied in younger adults, and Experiment 1 is designed to determine the extent to which older adults benefit from envelope cues for phonemic restoration. In addition, because phonemic restoration has only been studied in a quiet background, it remains unclear whether envelope cues contained in intervening noise will help or hinder speech recognition with a competing talker. For example, modulation masking from a competing talker may disrupt the benefit that envelope cues provide in a quiet background. Additional informational masking may result in listeners inappropriately linking speech segments from the competing talker to the target speech. Finally, older adults are known to be particularly sensitive to masking effects by competing speech, and a competing talker may compound the already challenging task of forming a coherent auditory object from interrupted speech. These research questions and hypotheses will be addressed in Experiment 1.

There are inconsistent findings regarding the potential advantage of periodicity cues for recognition of interrupted speech and phonemic restoration. In the absence of periodicity cues, recognition of vocoded sentences declines steeply when interrupted by silence (Jin & Nelson, 2010) and the addition of noise to the silent gaps does not typically result in phonemic restoration (Başkent, 2012; Başkent & Chatterjee, 2010; Chatterjee, Peredo, Nelson, & Başkent, 2010). However, the addition of periodicity information to vocoded speech improves recognition of interrupted vocoded speech and facilitates masking release (Başkent & Chatterjee, 2010; Stickney et al., 2007). Similarly, recognition of monaural interrupted vocoded speech improves when listeners are provided a continuous source of periodicity information in the opposite ear (Oh, Donaldson, and Kong, 2016). In contrast, a recent study using natural speech indicated that inconsistencies in periodicity information from glimpse to glimpse did not eliminate phonemic restoration when silent intervals where filled with noise (Clarke, Gaudrain, Chatterjee, and Başkent, 2014). This was interpreted as an indication that consistent voicing information across glimpses is not necessary for the object formation benefit associated with phonemic restoration. If continuous periodicity cues are beneficial for connecting glimpses of speech across time, then a non-speech filler signal that carries periodicity cues should benefit recognition of interrupted speech, and a filler signal that provides both envelope and periodicity cues should provide greater benefit than envelope cues alone. The advantage of continuous periodicity cues may be most apparent in a complex background, where listeners may rely on periodicity information to segregate speech from a competing talker. Finally, if sensitivity to periodicity cues declines with age, then older adults may receive less benefit than younger adults from continuous periodicity cues in quiet and/or competing talker backgrounds. These research questions and hypotheses will be addressed in Experiment 2.

I.C. Object Selection

Object selection refers to the process of choosing a particular auditory object to be the focus of attention and higher-level processing (Shinn-Cunningham, 2008). This process is often guided by *a priori* knowledge or expectations about the target, such as an expected spatial location, overall level relative to the background, and/or voice characteristics (Brungart, 2001; Kidd, Arbogast, Mason, & Gallun, 2005; Mackersie, Dewey, & Guthrie, 2011). These cues prime the listener to organize the auditory scene such that the appropriate object is represented in the foreground, with irrelevant competing signals as the background. Once an auditory object is selected by the listener, the neural representation of that sound source is enhanced relative to other competing sounds in the environment (Ding & Simon, 2012; Kerlin, Shahin, & Miller, 2010; Mesgarani & Chang, 2012). The contrast between auditory foreground and background allows the listener to identify and track the target over time, as well as avoid unwanted intrusions of irrelevant competing signals into higher cortical levels in the auditory system (Zion Golumbic et al., 2013). As the listener is usually responsible for deciding which object is the target, object selection involves a greater degree of intention than object formation, which is typically described as a lower level, automatic process (Bregman, 1990; Shinn-Cunningham, 2008). In a real-world environment with multiple talkers, younger listeners shift their attention seamlessly based on their intention and expectations of turn-taking in the conversation. Declines in the ability to quickly and efficiently switch the focus of attention to different voices may underlie age-related difficulty with speech recognition in realistic environments.

Whereas object selection and object formation can be viewed as distinct processes, the two likely work in tandem to facilitate perceptual organization of the auditory scene. Furthermore, the cues listeners use for object selection may influence object formation. For example, differences in talker sex may influence object formation based on F0 as a segregation cue, or selection based on listener expectations of the target’s voice, or both (Darwin, Brungart, & Simpson, 2003; Mackersie et al., 2011). Though object formation and selection are interrelated and mutually supportive, deficits in either may lead to difficulty understanding speech in a realistic background (Ihlefeld & Shinn-Cunningham, 2008). Evidence for age-related declines in both object formation and object selection can be found across several studies of speech recognition with competing talkers. For example, the finding that keyword recognition improves more over the course of a sentence for older adults than younger adults suggests that object formation may occur more slowly with increasing age (Ben-David, Tse, & Schneider, 2012; Ezzatian, Li, Pichora-Fuller, & Schneider, 2015). Other studies have observed that older adults are more likely to incorrectly repeat words from a competing talker than younger adults (Helfer & Freyman, 2008; Lee & Humes, 2012). This increase in “masker errors” suggests that object selection may be more difficult for older adults. Experiments that can evaluate object formation and selection separately are critical for assessing the relative contributions of these processes to speech perception difficulty in older adults. Experiments 1 and 2 are designed to assess effects of age on contributions of envelope and periodicity cues to object formation and an evaluation of keyword position effects will determine if object formation occurs more slowly for older adults. Experiment 3 assesses the effects of age and expectations of talker sex on object selection. Separate scoring for masker errors will help determine if object selection difficulties are greater among older adults.

I.C.1. Selection of speech objects

In a realistic conversation, the listener’s familiarity with the talker’s voice can facilitate object selection and improve speech recognition. Repeated exposure to a particular talker or set of talkers results in improved recognition of novel sentences spoken by those talkers (Yonan & Sommers, 2000) and fewer errors on a shadowing task (Newman & Evers, 2007). Importantly, these effects are observed in tasks with competing sources of speech, suggesting that talker familiarity may improve listeners' ability to organize a complex auditory scene and select the appropriate talker (Johnsrude et al., 2013; Newman & Evers, 2007). Whereas older adults are poorer than younger adults at identifying voices they have heard previously, they retain voice familiarity benefits on sentence recognition tasks (Johnsrude, Mackey, Hakyemez, Alexander, Trang, & Carlyon, 2013; Yonan & Sommers, 2000). These results suggest that older adults may use their knowledge and familiarity with a communication partner's voice to facilitate segregation and selection of target speech in realistic listening environments and potentially offset declines in object selection.

Most real-world environments allow listeners to generate expectations of a talker's voice based, minimally, on the talker's sex. Talker sex affects primarily two acoustic characteristics; fundamental frequency (F0), corresponding to the rate of vocal fold vibration and the perception of voice pitch, and the shape of the spectral envelope, corresponding to vocal tract length and the perception of voice timbre (Darwin et al., 2003). Male voices are characterized by a lower F0 and narrower spectral envelope than female voices; concurrent changes in F0 and spectral envelope can alter the perceived sex of a talker from male to female or vice versa (Darwin et al., 2003; Peterson & Barney, 1952). The acoustic and perceptual correlates of these cues are useful for object formation and object selection and many studies of speech recognition with competing talkers have used differences in sex or voice characteristics to help listeners identify the target (e.g., Duqesnoy, 1983; Festen & Plomp, 1990; Helfer & Freyman, 2008;

Mackersie et al., 2011). Additionally, poorer talker sex identification and speech recognition with a competing talker have been observed with vocoded speech, which is characterized by distortions of these important voice cues (Gaudrain & Carlyon, 2013; Gnansia, Pressnitzer, Pean, Meyer, & Lorenzi, 2010; Schvartz & Chatterjee, 2012). If older adults rely more on familiarity and expectations of a talker's voice to select the appropriate speech object in a complex background, they may be less able to shift their attentional focus when faced with unexpected changes in these voice characteristics.

I.C.2. Attentional filtering

The influence of attention on auditory perception has been demonstrated using non-speech signals modified to manipulate the listener's expectations (i.e., Scharf, Quigley, Aoki, Peachey, & Reeves, 1987; Schlauch & Hafter, 1991). The results of several studies using the "probe-signal method" have documented the shape and bandwidth of the attentional filter, a function describing the influence of focused attention on detection of expected vs. unexpected signals. The probe-signal method was first introduced by Greenberg and Larkin (1968) who reported that a listener's selection criteria for detection of a fixed-frequency tone in broadband noise resembles a simple band-pass filter. In their initial experiments, listeners were trained to detect a 1000 Hz tone (known as the "primary") in broadband noise using a two-alternative forced-choice design. In a small percentage of trials, the primary tone was replaced with a "probe" tone of the same intensity but at a frequency remote from the 1000 Hz primary. Listeners were not informed of the existence of the probes and were instructed in a way that would encourage focused attention for detection of the primary. Tone detection hovered around 80-90% correct across listeners for the 1000 Hz primary and declined to chance level for probes at remote frequencies. Their results illustrate that the listener’s detection of tones could be modeled as a band-pass filter, centered on the frequency of the primary, and the bandwidth of the filter was consistent with estimates of the critical band around 1000 Hz (Scharf, 1961).

Following this initial investigation, the probe-signal method has been used to describe other perceptual dimensions in which listeners can sharpen their attentional focus. These applications have incorporated a “cue” preceding the two observation intervals which identifies the primary frequency or signal to the listener (e.g., Dai, Scharf, & Buus, 1991; Scharf et al., 1987; Schlauch & Hafter, 1991). By collecting a complete psychometric function, data from the probe-signal method can be converted to a measure of decibel loss for unattended probes compared to expected primaries. Estimates of attenuation for probes falling outside the listening band vary from 3-7 dB across studies (Botte, 1995; Dai et al., 1991; Moore, Hafter, & Glasberg, 1996), with more recent work suggesting separate effects of attenuation for probes and enhancement of primaries, both of which combine to equal ~ 6 dB difference for detection of primaries compared to distant probes (Tan, Robertson, & Hammond, 2008). Other investigations have used probes with varying temporal features to show that listener expectations of signal duration and temporal structure influence signal detection in noise (Dai & Wright, 1995; White & Carlyon, 1997; Wright &

Dai, 1994). Thus, temporal integration may also be under some degree of attentional control, such that the size and position of integrative “multiple looks” (i.e., Viemeister & Wakefield, 1991) may depend on the listener’s expectations of the signal’s temporal structure. These results suggest that listeners may be able to use focused attention to modulate their sensitivity to spectral and temporal features of expected auditory targets.

Object selection in a multi-talker environment is driven by listeners' expectations of spectral and temporal features of the target. One of the simplest expectations for a speech object is the talker’s sex, based acoustically on F0 and the spectral envelope. By manipulating the listener’s expectations of the target voice, it may be possible to assess the extent to which focused attention to voice characteristics contributes to object selection in a two-talker environment. If the same tenets of attentional filtering described by the probe-signal method apply to selection of a target talker in a two-talker listening environment, then speech recognition should be best when listeners focus attention on voice features that match the target exactly. If the target’s voice features differ from the listener’s expectations, then speech recognition should decline. Effects of aging may be revealed by differences in peak performance, where focused attention facilitates object selection, and/or differences in the pattern of declines in speech recognition when the target’s voice features differ from the listener’s expectations. When listeners are unsure which talker is the target, they may be more likely to respond with keywords from the competing talker’s sentence. Quantifying masker errors separately from correct keywords may reveal age-related differences in the ability to accurately select the correct talker without the benefit of focused attention on voice features. These hypotheses will be tested in Experiment 3.

I.D. Cognitive Factors

The extent to which age-related declines in speech recognition can be explained by differences in cognitive abilities has been the focus of considerable research efforts (Humes et al., 1994; 2013; Füllgrabe et al., 2015; Rudner, Foo, Rönnberg, & Lunner, 2009; van Rooij, Plomp, & Orlebeke, 1989). Results of these studies have been mixed. As reviewed by Akeroyd (2008) in a meta-analysis, relationships are often found between cognitive and speech recognition measures, but no single cognitive test has been consistently linked to speech recognition. Lack of consistency is not entirely surprising, as differences in speech stimuli and cognitive tasks can make comparisons across studies difficult. Some of the most common cognitive abilities across studies include processing speed, working memory capacity, inhibitory control, and linguistic closure (Humes et al., 1994; 2013; Füllgrabe et al., 2015; Janse, 2012; Rudner et al., 2009). The relative contributions of these specific cognitive abilities to both object formation and object selection are of particular interest, as they may help explain the variance in speech recognition observed among younger and/or older adults. A cognitive battery was constructed to evaluate specific hypotheses related to age-related declines in cognitive abilities and their effects on speech recognition and perceptual organization.

I.D.1. Processing speed

Processing speed is a general term used to describe the speed or efficiency with which an individual can make a decision or perform a task. Age-related declines in processing speed are evident across a number of different tasks assessing both mental and manual processing speed (Salthouse, 2000). Speech naturally unfolds over time and adequate processing speed may be required to organize incoming speech information in an efficient manner. In the context of perceptual organization, it is hypothesized that processing speed plays a role in glimpsing, such that age-related declines in processing speed may limit speech recognition when shorter glimpses were presented at a fast rate. As processing speed can be expressed in both a mental and manual capacity, two measures were included in the cognitive battery. The Connections test (Salthouse et al., 2000) is a variant of the trail-making test designed to assess mental processing speed and the Purdue Peg Board test (Tiffin & Asher, 1948) is a measure of motor planning and manual speed of processing. While these two measures of processing speed have been shown to correlate within individuals, they capture distinct manifestations of age-related declines in processing speed and have been shown to predict performance on gap detection tasks (Harris, Eckert, Ahlstrom, & Dubno, 2010; Harris, Wilson, Eckert, & Dubno, 2012). It is unclear which of these two measures of processing speed would best predict speech recognition, and so both were included in the cognitive battery.

I.D.2. Working memory capacity

Working memory capacity refers to the amount of information that can be held in storage and retrieved during processing. The Reading Span Test (Daneman & Carpenter, 1980; Rönnberg, 1990) is a measure of working memory capacity that has been shown to predict speech recognition in older adults in multiple studies (Foo, Rudner, Rönnberg, & Lunner, 2007; Lunner, 2003; Rudner et al., 2009; Souza & Arehart, 2015). These findings supported a hypothesis that age-related declines in working memory contribute to poor speech recognition among older adults (Pichora-Fuller, Schneider, & Daneman, 1995). However, the relationship between working memory and speech recognition in normal hearing adults is still under debate (see Füllgrabe & Rosen, 2016; Gordon-Salant & Cole, 2016). In the context of perceptual organization, it is hypothesized that working memory capacity may affect the ability to hold glimpses of speech and restore missing segments to form a coherent object. The importance of working memory capacity may also depend on the length of the sentence, such that longer sentences may place greater demands on working memory than shorter sentences. Finally, sentences with a competing talker may result in unwanted intrusions of competing speech into working memory, which may have a more detrimental effect of performance for listeners with low working memory capacity. The Reading Span Test was included in the cognitive battery to test these hypotheses and determine if individual differences in working memory capacity influenced speech recognition, particularly with a competing talker and for sentences with a large number of keywords.

I.D.3. Inhibitory control

Inhibitory control refers to the ability to ignore or suppress task-irrelevant stimuli. In a competing talker background, the extent to which listeners can ignore competing speech may decline with age, resulting in unwanted intrusions of competing speech into the higher levels of processing (Humes, Lee, & Coughlin, 2006; Humes et al., 2013; Li, Daneman, Qi, & Schneider, 2004; Tun et al., 2002). Inhibitory control, as measured by the Stroop task (Stroop, 1935), may be sensitive to these potential declines in object selection. An essential element of the Stroop task is ignoring task-irrelevant linguistic information; the degree of interference measured on the Stroop task has been shown to predict distraction by competing speech in older adults on an auditory monitoring task (Janse, 2012). This suggests that age-related declines in inhibitory control may contribute to speech recognition difficulty of older adults with competing talkers; the Stroop task was included in the cognitive battery to test this hypothesis.

I.D.4. Linguistic closure

Linguistic closure refers to the ability to make use of partial linguistic information. In a realistic listening environment, glimpses of speech are inherently separated by gaps of missing information and the message must be inferred from the available glimpses. The ability to recognize a sentence based on partial information can be assessed in the visual domain with the Text Reception Threshold (Zekveld, George, Kramer, Goverts, & Houtgast, 2007). Performance on this measure has been shown to decline with age (see Humes, Kidd, & Lentz, 2016), as well as to predict recognition of speech in noise and interrupted speech (George, Zekveld, Kramer, Goverts, Festen, & Houtgast, 2007; Humes et al., 2013; Krull et al., 2013). This measure was included in the cognitive battery to test the hypothesis that an amodal ability to use partial linguistic information influences speech recognition in realistic listening environments.

I.E. Summary

Perceptual organization of speech in realistic listening environments is a complex process with several components. Difficulty with object formation or object selection will lead to the same basic outcome, that is, poorer speech recognition in a background of competing talkers. Older adults have particular difficulty understanding speech in realistic backgrounds and part of their difficulty may be age-related declines in one or more components of perceptual organization. The ability to use envelope and periodicity cues for object formation may decline with increasing age, and the relative contributions of these cues may differ depending on the complexity of the background. Age-related declines in object selection may manifest as poorer speech recognition due to difficulty using focused attention to select the appropriate talker. Additional factors may also contribute to perceptual organization, such as the lexical characteristics of the words in a sentence, or specific cognitive abilities of the listeners. All of these factors may independently contribute to aspects of speech recognition in realistic listening environments and each may be differentially affected by age. A series of experiments was designed to address these gaps in knowledge.

Three experiments were designed to determine the extent to which age-related declines in speech recognition could be explained by changes in several components of perceptual organization of speech. A general framework for perceptual organization and the three experiments is displayed in Figure 1. Experiments 1 and 2 determined the effects of age and the relative contributions of envelope and periodicity cues on three components of object formation: glimpsing, phonemic restoration, and speech segregation. Experiment 3 investigated the influence of focused attention and expectations of voice characteristics for object selection in a two-talker environment.

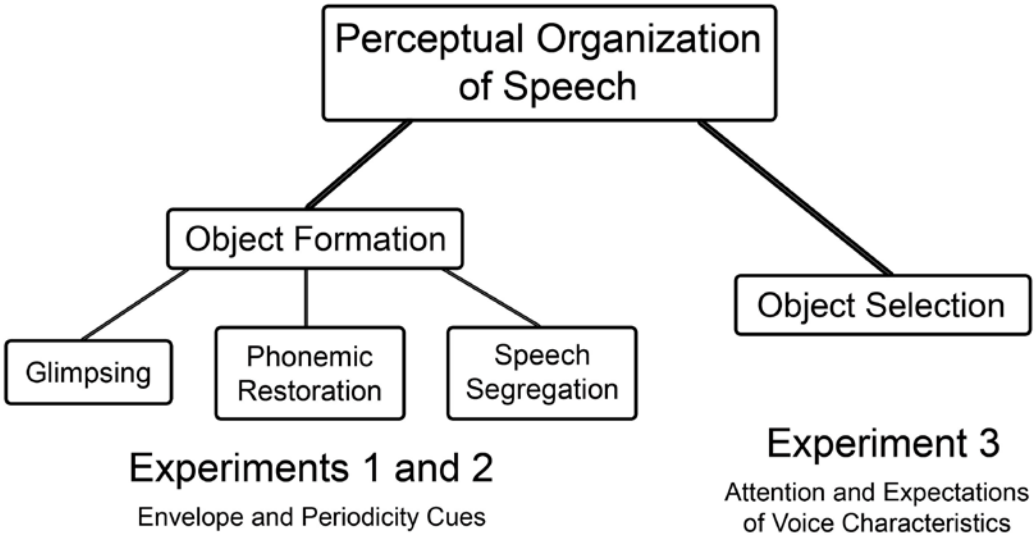

Figure 1: Perceptual organization of speech consists of two primary component processes, object formation and object selection. Object formation can be further described as the combined effects of glimpsing, phonemic restoration, and speech segregation. Experiments 1 and 2 investigated the relative contributions of envelope and periodicity cues to these three components of object formation. Experiment 3 investigated the role of attention and expectations of voice characteristics for object selection.

In each experiment, the contributions of several item-level factors were also evaluated. These included lexical factors, such as word frequency, neighborhood density, and biphone probability, as well as talker sex, the number of keywords in the sentence, and the relative position of the keyword within the sentence. Many of these item-level factors were found to differentially affect performance for younger and older adults and, in some cases, influence aspects of perceptual organization. Potential contributions of age-related declines in cognitive abilities and other subject-level factors were also explored across all 3 experiments. Subjects completed a cognitive test battery to assess processing speed, working memory capacity, inhibitory control, and linguistic closure. Subject-level differences in education level and pure-tone thresholds were also assessed as potential contributors to individual differences in speech recognition. In many cases, these subject-level factors were strongly associated with subject age, but analyses revealed a subset of subject-level factors that influenced speech recognition beyond effects of age. Conclusions that follow from these experiments support the hypothesis that age-related declines in perceptual organization of speech contribute to speech recognition difficulty of older adults in realistic listening environments.

II. Experiment 1: Effects of Age on Recognition of Speech Glimpses, Phonemic Restoration, and Speech Segregation

II.A. Introduction

In daily life, speech communication occurs in environments where background sounds fluctuate in level and mask portions of the intended message. Three distinct auditory processes may contribute to speech recognition in these environments; glimpsing, phonemic restoration, and speech segregation. “Glimpsing” refers to the process of identifying audible fragments of speech and connecting them together across gaps of missing information to form a single coherent stream (Cooke, 2006; Moore, 2003). “Phonemic restoration,” refers to the process of filling in missing information based on the available acoustic information, knowledge of the language, and semantic context (Bashford & Warren, 1987; Warren, 1970). When the background includes other talkers, “speech segregation” is the process by which glimpses from the talker of interest are perceptually separated from other sources of speech so that attention can be focused on a single target (Brungart, 2001; Shinn-Cunningham, 2008). Relative to younger adults, older adults require a more advantageous SNR in a fluctuating background and they are more likely to mistake speech from a competing talker for the intended message (Festen & Plomp, 1990; Helfer & Freyman, 2008; Lee & Humes, 2012). Although differences in hearing sensitivity can contribute to (or compound) the effects of age on speech recognition, age-related declines in speech recognition have been demonstrated in older adults with normal hearing and under conditions in which differences in audibility are minimized (Dubno et al., 2002; 2003; Eisenberg, et al., 1995; Rajan & Cainer, 2008). Recent research has identified several factors other than hearing sensitivity that may contribute to this age-related difficulty, including declines in specific cognitive abilities and poor auditory temporal processing (Füllgrabe et al., 2014; Humes et al., 2013). However, these studies concluded that the combined effects of these factors and hearing sensitivity accounted for only about 60% of the variance in speech recognition among older adults in realistic listening environments (Füllgrabe et al., 2014; Humes et al., 2013), leaving a considerable amount of residual variance unexplained. The purpose of Experiment 1 was to determine the extent to which this residual variance in speech recognition among older adults with normal hearing may be explained by declines in glimpsing, phonemic restoration, and/or speech segregation.

Glimpsing has been studied using various experimental methods in younger and older adults. One method has been to compare speech recognition in steady-state noise to performance in a gated or modulated noise (see Moore, 1990 for a review). Typical Gaussian noise contains random amplitude fluctuations that result in modulation masking of important envelope cues in speech (Stone et al., 2012). Momentary reductions in the level of the noise allow listeners to glimpse speech at favorable SNRs without modulation masking, which results in improved speech recognition compared to the steady-state masker. Older adults typically benefit less from glimpsing as compared to younger adults (Festen & Plomp, 1990; Eisenberg et al., 1995; Takahashi & Bacon, 1992). However, forward masking may limit recognition of sentences in interrupted noise, particularly for older adults (Dubno et al., 2002; 2003). Glimpsing can be studied in the absence of forward masking by replacing portions of a sentence or word with silence (Gilbert et al., 2007; Jin & Nelson, 2006; 2010; Miller & Licklider, 1950; Wang & Humes, 2010). Recognition of silence-interrupted speech is determined primarily by the proportion of speech information that remains after interruption and is available for glimpsing (Kidd & Humes, 2012; Wang & Humes, 2010). For a given proportion of speech, age-related declines in recognition are typically observed for slower interruption rates, between 2-5 Hz (Saija et al., 2014; Shafiro, Sheft, Risley, & Gygi, 2016). These rates approximate the syllabic rate of speech, and would be representative of the interruptions imposed by a competing talker. A common finding across these studies is that hearing loss contributes strongly to recognition of glimpsed speech, but that residual effects of age persist even in the absence of hearing loss (Dubno et al., 2003; Krull et al., 2013; Molis, Kampel, McMillan, Gallun, Dann, & Konrad-Martin, 2015; Shafiro et al., 2016). Taken together, age-related declines in connecting audible portions of speech across time may contribute to speech recognition difficulty of older adults in realistic listening environments.

Glimpsed speech inherently includes missing information, which must be perceptually restored to facilitate speech recognition. This process can be studied with the phonemic restoration paradigm, where noise bursts are inserted into the silent intervals of interrupted speech (Bashford & Warren, 1987; Başkent, Eiler, & Edwards, 2009; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983). The addition of noise to silent intervals is interpreted by the auditory system as evidence that the speech signal is continuous, but portions are being masked by the relatively high-level noise bursts (typically -10 dB re: speech glimpses). The illusion of continuity ("picket fence percept," Miller & Licklider, 1950) results in a more coherent perception of the glimpsed speech signal as well as improved speech recognition (Bregman et al., 1999). Phonemic restoration can be enhanced when intervening noise is modulated by the envelope of the missing speech segment (Bashford et al., 1996; Shinn-Cunningham & Wang, 2008). This finding suggests that continuity of envelope cues over the course of the sentence can facilitate glimpsing and accurate restoration of missing speech segments. This effect has only been studied in younger adults and so the extent to which older adults can use envelope cues to restore missing speech information remains unclear.

Relatively little is known about the effects of age on phonemic restoration. Some studies have shown minimal phonemic restoration in older adults (Başkent et al., 2009; Madix, Thelin, Plyler, Hedrick, & Malone, 2005). However, the design of these studies was such that the investigators were unable to disentangle effects of age from those of peripheral hearing loss. Only one study to date has investigated phonemic restoration in older adults with normal hearing and those results suggested an *improvement* in phonemic restoration with age (Saija et al., 2014). However, this finding was limited to a subset of conditions with a relatively slow interruption rate (2.5 Hz). In addition, the relatively high-level noise bursts that are typically used to elicit phonemic restoration may result in forward masking which has been shown to disproportionately limit speech recognition for older adults (Başkent et al., 2009; Bashford, Riener, & Warren, 1992; Dubno et al., 2002; 2003). The use of envelope-modulated noise allows phonemic restoration to be elicited with a lower level noise (0 dB SNR; Shinn-Cunningham & Wang, 2008). This may limit the confounding effect of forward masking in measures of phonemic restoration.

Language abilities and use of context are known to remain stable with increasing age (e.g., Pichora-Fuller, 2008; Salthouse, 2010) and these abilities may contribute to phonemic restoration among older adults. Lexical characteristics of keywords, such as word frequency, neighborhood density, and biphone probability, provide a means to compare phonemic restoration between lexically easy and lexically difficult words. If phonemic restoration depends on language abilities, then restoration should help listeners accurately restore lexically difficult words that would otherwise be unintelligible when sentences are interrupted with silence. This hypothesis predicts greater phonemic restoration for words that are used less frequently, have more lexical neighbors, and have less common phonotactic patterns. If older adults are able to leverage their strong language abilities for phonemic restoration, then they should demonstrate greater phonemic restoration than younger adults for the more lexically difficult keywords in sentences. This hypothesis was tested in Experiment 1 using an item-level analysis of phonemic restoration based on the lexical characteristics of the keywords.

Segregation of speech from a competing talker may also become more difficult with age and may place additional demands on glimpsing and phonemic restoration. Many studies have demonstrated an age-related decline in speech recognition with a competing talker, but identifying the underlying source of this decline has proven difficult. A competing talker background results in a combination of masking effects, including energetic masking, which may separate target speech into a series of glimpses, and informational masking, which requires segregation and focused attention to identify target speech and ignore the competing message. As such, age-related declines in speech recognition with a competing talker have been interpreted in several different ways. For example, age-related declines in speech recognition have been associated with an increase in “masker errors,” where subjects incorrectly respond with words from the masker sentence (Helfer & Freyman, 2008; Humes et al., 2006). These findings provide support for a hypothesis that speech recognition difficulty stems from age-related declines in higher-level functions, such as attention to target speech, and/or inhibition of competing speech (Janse, 2012; Tun et al., 2002). Other studies have shown that speech recognition with a competing talker improves over the time course of the sentence, suggesting that speech segregation may occur more slowly for older adults (Ben-David et al., 2012; Ezzatian et al., 2015). Considering the evidence that older adults are poorer at recognizing glimpsed speech, even in the absence of a competing talker, adding the perceptual demands associated with speech segregation and focused attention may compound the effects of glimpsing and further erode performance by older adults. This hypothesis was tested in Experiment 1 using an experimental design in which glimpsing and speech segregation were assessed independently.

Several studies have suggested that age-related declines in cognition may contribute to poorer speech recognition among older adults (Füllgrabe et al., 2015; Humes et al., 1994; 2013; Pichora-Fuller et al., 1995). Whereas relationships between cognitive measures and speech recognition abilities are often observed, results across studies lack replication (Akeroyd, 2008). Some degree of inconsistency across studies is likely related to differences in choice of speech stimuli (i.e., words vs. sentences) and background conditions (noise vs. speech), which likely influence the relative contributions of different cognitive abilities. In addition, cognitive demands may differ for speech recognition tested at or near threshold, where many acoustic cues are degraded or inaudible, as compared to perception of stimuli presented at suprathreshold levels. In most cases where effects of cognition on speech recognition are observed, the magnitude of the effects are small, which may also contribute to the inconsistencies across studies. The approach taken in Experiments 1 and 2 was to use a battery of cognitive measures to assess several dimensions of cognition, including processing speed, working memory capacity, inhibitory control, and linguistic closure. Cognitive tests were selected based on previous research identifying relationships between those measures and glimpsing, phonemic restoration, and/or speech segregation. These relationships were explored using statistical modeling to account for differences between younger and older adults as well as collinearity among cognitive abilities.

The goals of Experiment 1 were to determine the extent to which aging affects (1) glimpsing, (2) phonemic restoration, and (3) speech segregation. Hypotheses predicted that poorer speech recognition of older adults would be partially explained by age-related declines in glimpsing and speech segregation, whereas older adults would benefit more than younger adults from phonemic restoration. Younger and older adults with normal hearing listened to sentences interrupted with either silence or envelope-modulated noise in quiet or with a competing talker. The proportion of the sentence remaining after interruption was manipulated to determine the extent to which glimpsing contributed to speech recognition; performance was expected to improve with increasing proportion of speech. Performance for sentences interrupted with envelope-modulated noise was expected to be better than for silence-interrupted sentences, and the magnitude of this improvement was defined as phonemic restoration. Finally, performance was expected to decline with a competing talker, and the extent of this decline reflects the added perceptual demands associated with speech segregation. Interactions among these factors revealed independent age-related changes in each of these abilities. Additional contributions from item-level factors (lexical characteristics, sentence length and keyword position, and talker sex) and subject-level factors (education level, hearing sensitivity, and cognitive abilities) were quantified within a statistical model to explore potential interactions with glimpsing, phonemic restoration, speech segregation, and age. The combined effects of these factors revealed the extent to which they explain age-related declines in speech recognition in realistic listening environments.

II.B. Methods

II.B.1. Subjects

A sample size of 40 subjects was chosen based on a power analysis of pilot data. With a power of 0.80, hypothesized effects of speech proportion, interruption type, background, and age could be expected with a significance level <0.001. Two groups were tested, including 20 younger adults ranging in age from 18 to 29 years (mean: 24.7, SD: 2.8), and 20 older adults ranging in age from 63 to 84 years (mean: 69.9, SD: 5.7). Older subjects were screened for normal cognitive functioning using the Mini Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975) and all subjects passed this screening with a score of 25 or greater (Tombaugh & McIntyre, 1992). All subjects were native speakers of American English.

Hearing sensitivity was assessed in all subjects based on air-conduction thresholds at audiometric frequencies (ANSI, 2010). Hearing threshold criteria for younger subjects was defined as thresholds ≤ 25 dB HL for 250-8000 Hz. For older subjects, threshold criteria were increased to ≤ 30 dB HL for 250-6000 Hz to ensure that adequate number of subjects could be recruited for testing. Mean audiograms for younger and older subjects are displayed in Figure 2. Although all subjects met hearing threshold criteria for participation, differences in mean thresholds were noted between younger and older subject groups, particularly at higher frequencies. To rule out the possibility that slightly elevated hearing thresholds for older adults contributed to their poorer speech recognition, several pure-tone averages (PTA) were calculated to quantify mean hearing sensitivity across different frequency ranges, based on work by Simpson, Matthews, and Dubno (2013); narrow PTA (0.5, 1.0, 2.0, 4.0 kHz), broad PTA (0.5, 1.0, 2.0, 3.0, 4.0, 6.0, 8.0 kHz), low-frequency PTA (0.25, 0.5, 1.0 kHz), and high-frequency PTA (2.0, 3.0, 4.0, 6.0, 8.0 kHz). Effects of hearing sensitivity on speech recognition were examined statistically using these PTA measures as potential contributing factors in a generalized linear mixed model.

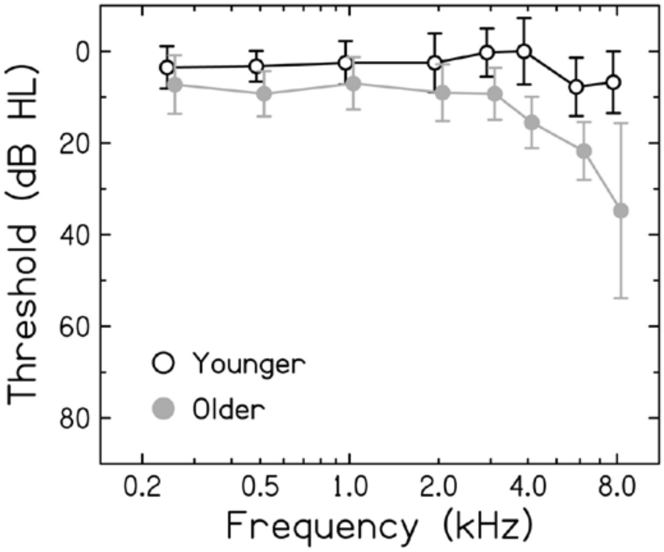

Figure 2: Mean audiograms for younger (open symbols) and older adults (gray). Error bars represent standard deviation. Narrow pure-tone average (NPTA) was 2.1 dB HL for younger adults and 10.2 dB HL for older adults. Broad pure-tone average (BPTA) was 3.3 dB HL for younger adults and 15.21 dB HL for older adults. Low-frequency pure-tone average (LFPTA) was 3.1 dB HL for younger adults and 7.8 dB HL for older adults. High-frequency pure-tone average (HFPTA) was 3.5 dB HL for younger adults and 18.1 dB HL for older adults.

II.B.2. Stimuli and apparatus

Speech stimuli were sentences from the Perceptually Robust English Sentence Test Open-set (PRESTO; Gilbert, Tamati, & Pisoni, 2013). PRESTO consists of 20 sentence lists, each containing 18 sentences spoken by different talkers from various dialect regions in the United States (TIMIT corpus; Garofolo et al., 1993). Each list contains 9 different male and female talkers, making the corpus well suited for analyzing talker sex as a potential factor affecting keyword recognition. The sentences vary in length, and contain from 3-6 keywords, for a total of 76 keywords per list. A recent investigation of list equivalency for PRESTO revealed only a subset of PRESTO lists were equivalent for certain types of speech processing and background conditions (Faulkner, Tamati, Gilbert, & Pisoni, 2015). Results of Faulkner et al. (2015) and pilot data on silence-interrupted and envelope-filled sentences were used to identify equivalent sentence list pairs for each speech proportion. Within these sentence list pairs, processing type was counterbalanced across subjects to ensure that the critical comparisons for investigating phonemic restoration were made between equivalent sentence lists and the type of processing applied to each list did not confound the result.

Interrupted versions of the sentences were generated by multiplying the original waveforms by a rectangular gating function with 10-ms transition ramps. Five gating functions were designed to generate interrupted sentences retaining proportions of speech of 0.40, 0.50, 0.55, 0.60, and 0.70. The proportion of speech corresponds to the duty cycle of the gating function (i.e., 0.50 speech proportion = 50% duty cycle), and the interruption rate was adjusted for each speech proportion to maintain 200-ms interruptions (Figure 3). A consistent duration of interrupted segments was desirable so that phonemic restoration of equal duration segments could be compared across the different proportions of speech. The interruption duration of 200-ms was selected so that interruptions would be less than the average word length, which has been shown to produce robust phonemic restoration (Bashford & Warren, 1987; Bashford et al., 1988). The gating function was anchored to the beginning of the sentence and began with a positive phase, such that the initial portion of the sentence was always gated on. The gating function repeated in a periodic fashion thereafter and ended with the offset of the sentence, often resulting in an incomplete final cycle of the gating function at the end of the sentence. As such, the exact proportion of speech for a given sentence varied slightly from the value assigned by the gating function.

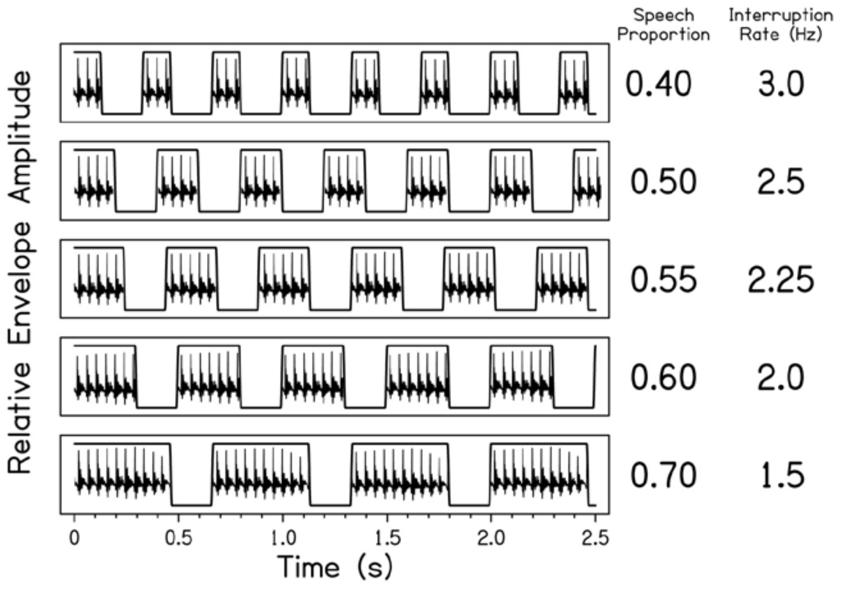

Figure 3: Schematic illustration of 5 gating functions used to interrupt sentences. The y-axis is the relative envelope amplitude, and the x-axis is time in seconds. Position of speech glimpses are indicated by waveform illustrations. Interruption rate and duty cycle of the gating functions were adjusted in tandem to maintain 200-ms interruptions for each speech proportion.

Original sentence waveforms were multiplied by the gating function to produce sentences interrupted by silence (Figure 4; green). For sentences filled with envelope-modulated noise, speech shaped noise was generated by filtering white noise by the long-term average speech spectrum (LTASS) of the TIMIT sentences. The LTASS was measured using an add-overlap method (e.g., Versfeld, Daalder, Festen, & Houtgast, 2000) with 16-ms Hanning windows and 50% overlap, measured over 18 concatenated sentences from PRESTO list 1. The envelope of a given sentence was obtained via low-pass filtering at 50 Hz and full wave rectification and then used to modulate the speech-shaped noise to generate envelope-modulated noise. The envelope-modulated noise was then amplified/attenuated to match the root-mean-square (RMS) amplitude of the original sentence and then multiplied by the inverse of the gating function, such that the noise is on during portions when the speech is off (Figure 4; black). The sum of the interrupted speech and noise is an interrupted sentence with envelope-modulated noise filling the interruptions at a 0 dB SNR. The resulting waveform closely matches the temporal waveform of the original speech signal (Figure 4).

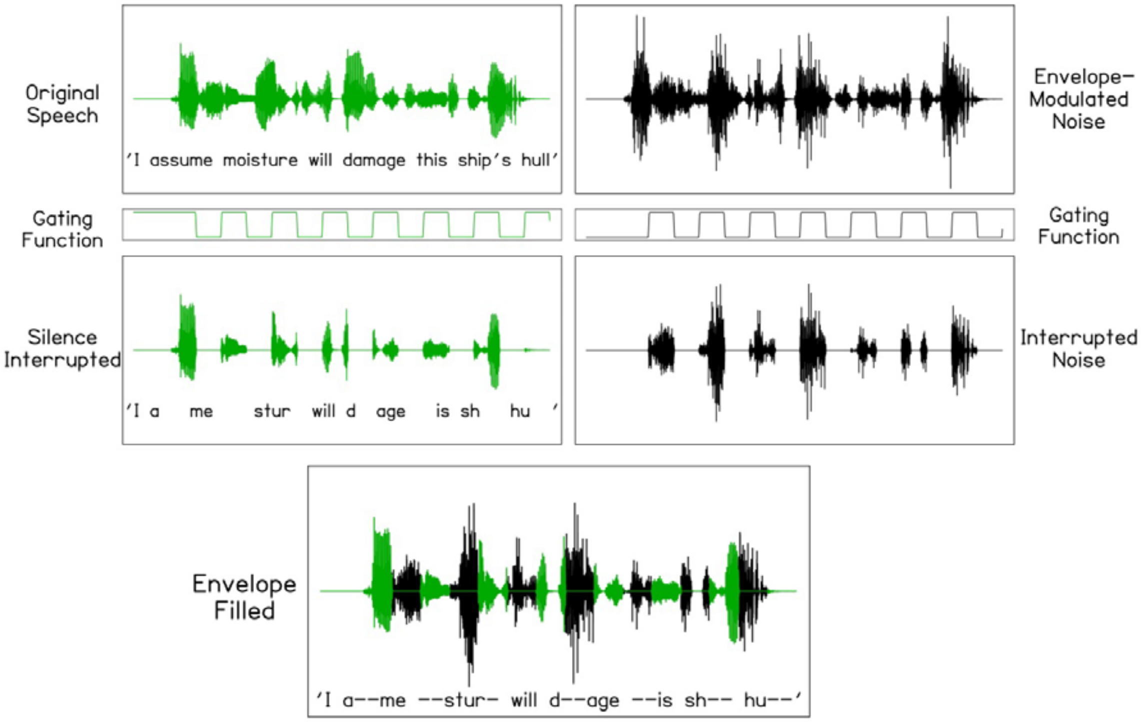

Figure 4: Schematic diagram of stimulus generation. Speech (green) and envelope-modulated noise (black) are multiplied by inverse gating functions and combined to create envelope-filled sentences with the same temporal envelope as the original speech.

For conditions with a competing talker, sentences from the TIMIT corpus (Garofolo et al., 1993) that were not used in any PRESTO list served as competing talkers. Competing talker sentences were chosen individually to match the duration of a corresponding target sentence from PRESTO. In addition, competing talker sentences were chosen such that the sex of the competing talker differed from the sex of the target talker. Unprocessed competing talker sentences were mixed with interrupted or envelope-filled target sentences at +3 dB SNR. Pilot data indicated that the 0.40 speech proportion with a competing talker resulted in floor performance and, as a result, this condition was dropped, leaving four speech proportion conditions; 0.50, 0.55, 0.60, and 0.70. To help subjects identify the target talker, each competing talker sentence mixture was preceded by a cue phrase spoken by the target talker in quiet. This cue phrase, “greasy wash water all year,” is a portion of a standard sentence recorded by each talker in the TIMIT corpus. Similar voice cues have been used to identify the target talker in other multi-talker speech tasks, such as the Coordinate Response Measure (Brungart, 2001) and the Theo-Victor-Michael sentences (Helfer & Freyman, 2009).

Sentences were pre-processed and saved as separate .wav files with 16-bit resolution at a sampling rate of 16000 Hz. A calibration noise with the same spectral shape and RMS amplitude as the uninterrupted sentences was set to 70 dB SPL through the headphones using an acoustic coupler with a Larson Davis model 2559 ½ inch microphone and a Larson Davis Model 824 sound level meter with flat weighting. The overall level of the interrupted sentences was expected to be somewhat less than 70 dB SPL and to vary from sentence to sentence based on the position of the interruptions; this method of calibration ensures that the “on” portions of the sentences are presented at 70 dB SPL. The overall level for stimuli with a competing talker was expected to be somewhat greater, due to the addition of the uninterrupted competing talker at a +3 dB SNR.

II.B.3. Procedures

II.B.3.a. Speech measures

Speech recognition testing was completed in a sound-attenuating booth in a single two-hour session. Presentation and keyword scoring was controlled by Token software (Kwon, 2012). Speech stimuli were output via computer with a Lynx Two multichannel audio interface through a Tucker-Davis Technologies programmable attenuator and headphone buffer and were presented monaurally through Sennheiser HDA 200 headphones. The right ear was chosen as the default test ear unless that ear did not meet hearing criteria; four older subjects were tested using the left ear for this reason. Testing was blocked by background (quiet/competing talker) and counterbalanced across subjects. Data collection in each background consisted of one list per interruption type (silence-interrupted or envelope-filled) at each speech proportion in a random order, for a total of 10 lists in the quiet background (5 speech proportions) and 8 lists in the competing talker background (4 speech proportions). A break was offered between quiet and competing talker test blocks.

Prior to each test block, subjects completed two practice lists designed for familiarization with the stimuli and response paradigm. Pilot data suggested that two lists were sufficient to minimize learning effects over the course of a block of testing. For testing in the quiet background, subjects were instructed to repeat each sentence, guessing whenever possible even if they were not able to understand all of the words in the sentence. For testing in the competing talker background, subjects were told they would hear a voice say the phrase “greasy wash water all year," followed by a mixture of two people talking simultaneously. They were instructed to use the cue phrase to help them identify the interrupted (target) sentence and repeat it aloud, guessing whenever possible. Subject responses were recorded using a Realistic Highball Dynamic 33-984C microphone, routed through the soundcard of a separate computer. Responses were scored live by the experimenter using a strict scoring rule (i.e., no additional or missing suffixes) and also recorded using Adobe Audition and saved as .mp3 files for offline review and confirmation of scoring, as needed.

II.B.3.b. Lexical characteristics

Lexical characteristics for the keywords were obtained using The Irvine Phonotactic Online Dictionary (IPhOD version 2.0; Vaden, Halpin, & Hickok, 2009). The IPhOD uses the SUBTLEXus database (Brysbaert & New, 2009) to generate context-weighted measures of word frequency, neighborhood density, and biphone probability. Word frequency measures were raw counts of the occurrence of a word in film subtitles, adjusted for the number of different films in which the word occurs (i.e., context-weighted, Brysbaert & New, 2009). For keywords with multiple pronunciation entries in the IPhOD, the entry that most closely matched the talker's production of the stimulus was selected. IPhOD entries were available for 98.5% of the keywords in the PRESTO corpus. Remaining keywords fit into four general categories; hyphenated words, proper nouns, monophonemic words, and uncommon words. The PRESTO corpus contains 15 hyphenated keywords (e.g., "part-time," "long-term") which account for 1.0% of total keywords in the corpus. Each hyphenated keyword consisted of two words with individual entries in the IPhOD. Lexical values for hyphenated keywords were calculated as the average word frequency, neighborhood density, and biphone probabilities of the two words contained within the hyphenated word. Three proper nouns in the corpus did not have word frequency counts in the SUBTLEXus; “Rachel,” “Greg,” and “Gwen.” As these items represented a very small portion of total keywords (0.2%), they were excluded from analyses that included word frequency as a factor. Similarly, the monophonemic keyword, “oh,” was used twice in the corpus (0.1% of total keywords) and these items were excluded from analyses that included biphone probability as a factor. The remaining two keywords without frequency counts were exceedingly uncommon words, “micrometeorite” and “unauthentic.” These items were assigned word frequency counts of 0, and measures of neighborhood density and biphone probability were calculated using phonetic transcription into the IPhOD. Lexical values were normalized across keywords to have a mean of 0 and standard deviation of 1 prior to statistical analyses.

II.B.3.c. Additional item-level factors

Three additional item-level factors were evaluated for each keyword; talker sex, sentence length, and keyword position. Each keyword in a given sentence was assigned values for the sex of the talker (-1 for male, +1 for female) and the total number of keywords in the sentence ranged from 3-6 across sentences. Keyword position was calculated as the keyword’s serial order position among keywords in the sentence, divided by the total number of keywords in the sentence. Sentence length and keyword position were each normalized across keywords so that both factors would have a mean of 0 and standard deviation of 1 prior to statistical analyses. This normalization procedure ensured that modeling results could be compared across different factors in a standardized scale.

II.B.3.d. Cognitive measures

In a separate two-hour test session, subjects completed a battery of cognitive measures designed to assess specific cognitive abilities that may contribute to speech recognition in realistic listening environments. The test battery consisted of measures of processing speed (Connections, Purdue Peg Board), working memory capacity (Reading Span), inhibitory control (Stroop), and visual linguistic closure (Text Reception Threshold). All subjects except one completed each cognitive measure. The one exception, a younger subject, reported blue-green colorblindness and was unable to distinguish blue from green text on the Stroop test form. The Stroop was not completed for this subject and that subject's data were excluded from analyses that included Stroop as a factor.

Connections: The Connections test (Salthouse et al., 2000) is a variant of the trail-making test designed to assess cognitive speed of processing. Each page of the test contains a number of circles containing letters and/or numbers, and subjects trace a line through the circles in a specified order. Four "simple" trials contained either all numbers or all letters, and subjects connect the circles in numerical or alphabetical order. Four "complex" trials contained both letters and numbers, and subjects connect circles in alternating numerical-alphabetical order. Each trial was scored for the number of circles correctly connected in 20 seconds. The mean of the 4 complex trials was recorded as the final score.

Purdue Peg Board: The Purdue Peg Board test (Tiffin & Asher, 1948) is a test of manual speed of processing. In the 3 “simple” conditions, subjects have 30 seconds to insert as many pegs into the board as possible using their right hand, left hand, and both hands. In the “complex” condition, subjects have 60 seconds to complete as many “assemblies” as possible; assemblies consist of a peg, a washer, a collar, and another washer, inserted into the board using both hands in an alternating fashion. Conditions are scored based on the total number of pieces correctly inserted into the board in the specified time. The mean of 2 repetitions of the complex condition was recorded as the final score.

Reading Span: The Reading Span Test (Daneman & Carpenter, 1980; Rönnberg, 1990) is a measure of working memory capacity. Sentences are presented on a computer screen, one or two words at a time, and subjects read the sentence aloud and report whether the sentence was semantically correct or incorrect. After a number of sentences are presented, subjects are instructed to recall either the first or last word in each sentence. First word/last word recall is assigned randomly for each sentence set and is unknown to the subject until the end of the set. The number of sentences in each set increases over the course of the test, from 3 to 6 sentences. The final score is the percentage of correctly recalled words across all 54 sentences in the test, excluding the initial training set.

Stroop: The Stroop Neuropsychological Screening Test (Trenerry, Crosson, Deboe, & Leber, 1989) was used as a measure of inhibitory control. The 2 trials of the test each consist of a page of 112 color words (blue, green, tan, red) printed in incongruent colored ink. On the first trial, subjects read the words aloud, ignoring the color that the word is printed in. On the second trial, subjects name the color of ink that each word was printed in, ignoring the word itself. Trials concluded after 120 seconds, or when the subject completed the page, whichever occurred first. The number of seconds elapsed was divided by the total number of correct responses to yield reaction times (seconds/word) for both conditions. Typically, Stroop interference is quantified as the difference in reaction time for reading the color words vs. naming the color of the ink. However, this metric can be confounded by differences in baseline word reading speed between younger and older listeners, and therefore differences in reaction time were normalized to the word reading condition to yield a final score (Davidson, Zacks, & Williams, 2003; Dulaney & Rogers, 1994; Hartley, 1993).

Text Reception Threshold: The Text Reception Threshold (Zekveld, George, Kramer, Goverts, & Houtgast, 2007) is a measure of visual linguistic closure. The English version of the test uses high-predictability sentences from the revised Speech in Noise (R-SPIN) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). On each trial, words from the R-SPIN sentences appear on a computer screen behind an array of vertical black bars. Subjects read as much of the sentence aloud as possible and are encouraged to guess if they are unsure. Sentences are scored as correct only if the entire sentence is repeated correctly. The width of the black bars, which determines the proportion of the sentence that is visible, varies adaptively based on the subject's performance. After 13 sentences are presented, a threshold estimate is calculated from the average proportion of visible text on trials 5-13. Four threshold estimates were obtained from four different sets of R-SPIN sentences and the mean of the four thresholds was recorded as the final score.

II.B.4. Statistical approach

Data were analyzed using an item-level logistic regression analysis of keyword recognition implemented in R (R Core Team, 2016) using a generalized linear mixed model (GLMM; lme4 software package; Bates, Maechler, Bolker, & Walker, 2015). Generalized linear mixed modeling yields separate β coefficients for the magnitude of fixed effects for each factor in the model (standard estimates). Item-level data were fit to the model to predict the binary dependent variable, keyword recognition (*W*). The GLMM analysis was performed for each keyword across all subjects with the following design-level factors: speech proportion (*Prop*), background (*Bg*), interruption type (*Int*), age group (*Age*), and random subject effects (*Subj*). Speech proportion values (0.40-0.70) were normalized for a mean of 0 and standard deviation of 1. Remaining predictor variables (*Bg*, *Int*, and *Age*) were binary in nature; positive sign indicated competing talker background (*Bg*), envelope-filled sentence (*Int*), and older adult (*Age*). The random subjects effects term included independent contributions of *Prop*, *Bg*, and *Int*, such that the model could adjust for subject-level differences in the effects of these factors (i.e., Clark, 1973). A simplified version of the model is expressed below.

$$W \sim Prop + Bg + Int + Age + (Prop + Bg + Int \mid Subj)$$

Several additional factors were evaluated using model testing (Hofmann, 1997) to determine whether the fit of the GLMM significantly improved with the addition of a given factor. These factors included interactions between two or more factors, item-level factors such as talker sex, sentence length, keyword position, and lexical characteristic of the keywords, and subject-level factors such as education level (in years), PTA measure, and cognitive test scores. A combination of step-wise factor addition and elimination was performed to optimize model fit with all significant factors and interactions. Some degree of correlation between fixed effects was expected (such as between various lexical factors for a given keyword). Such correlations can challenge the interpretability of the results, as the effects of one factor are likely to influence the model’s ability to fit β values to the other factor. A correlation of fixed effects table (i.e., covariance matrix) was constructed to check for collinearity. Correlations between fixed effects that exceeded 0.3 (i.e., greater than negligible correlation; Hinkle, Wiersma, & Jurs, 2003) were addressed by residualizing and rescaling predictor variables to reduce collinearity. To residualize a given factor (*A*) by a collinear factor (*B*), a general linear model is constructed in which factor *A* is predicted by factor *B* ($A \sim B$). The residuals from this model represent the variance in factor *A* that cannot be explained by factor *B*. These model residuals are uncorrelated with factor *B* and can be normalized and used in the GLMM in place of the original factor *A*.

A single model was constructed, but modeling results will be discussed in three sections for clarity: design-level factors, item-level factors, and subject-level factors. Significant interactions were interpreted using separate *post-hoc* models with split factors. Separate *post-hoc* models were constructed for each interaction to describe the effect of a given factor across the two levels of its interacting factor. For a given interaction (Factor A × Factor B), the values of Factor A were separated into two factors, one for each level of the binary interacting factor (Factor B). These two “split factors” replaced Factor A and the A×B interaction term in the original model. The relative magnitudes of standard estimates for the split factors in the resulting model indicated the nature of the interaction in the original model. Standard estimates, standard error, and *z*-statistics for each factor, including split factors from *post-hoc* models, can be found in Appendix 1.

II.C. Results

II.C.1. Design-level factors

Keyword recognition, transformed into rationalized arcsine units (rau; Studebaker, 1985), is plotted in the top panels of Figure 5 as a function of proportion of speech for younger (left) and older adults (right) with interruption type and background as parameters. Phonemic restoration was defined as the difference in keyword recognition for silence-interrupted and envelope-filled sentences and is plotted as histograms in the bottom panels of Figure 5.

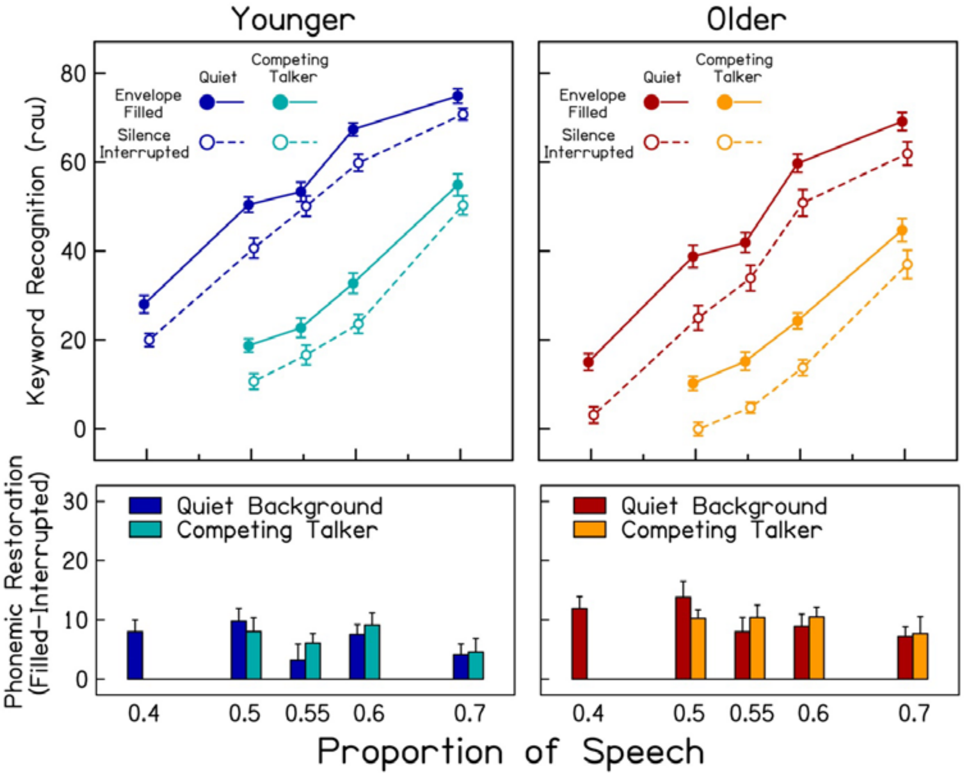

Figure 5: Top panels: Keyword recognition (rau) plotted as a function of proportion of speech with interruption type and background environment as parameters. Younger adults are plotted on the left, and older adults are plotted on the right. Bottom panels: Phonemic restoration, quantified as the difference in keyword recognition between silence-interrupted and envelope-filled sentences, is plotted as a function of proportion of speech with background environment as the parameter.

A GLMM was constructed to quantify the relative contributions of the design-level factors and their interactions to keyword recognition. Model testing confirmed significant improvements in the fit of the model with the addition of each design-level factor (χ^2_{Prop}=181.88, $p \leq 0.001$; χ^2_{Bg}=151.95, $p \leq 0.001$;

χ^2_{Int}=76.05, $p \leq 0.001$; χ^2_{Age}=20.78, $p \leq 0.001$). Modeling results indicated that younger adults significantly outperformed older adults (β_{Age}=-0.29; z=-5.23; $p \leq 0.001$). Keyword recognition improved significantly with increasing proportion of speech (β_{Prop}=1.03; z=58.41; $p \leq 0.001$), indicating that the amount of glimpsed speech was a strong predictor of performance. Keyword recognition also improved significantly with the addition of envelope-modulated noise to silent intervals (β_{Int}=0.23; z=14.51; $p \leq 0.001$), indicating that phonemic restoration facilitated recognition of envelope-filled sentences. Keyword recognition declined significantly with the addition of a competing talker (β_{Bg}=-0.75; z=-39.11; $p \leq 0.001$), indicating a considerable increase in task demands associated with segregating speech glimpses from a competing talker. Several interactions between the design-level factors also improved the fit of the model ($\chi^2_{Prop*Bg}$=43.99, $p \leq 0.001$; $\chi^2_{Prop*Age}$=15.91, $p \leq 0.001$; $\chi^2_{Prop*Int}$=29.40, $p \leq 0.001$; $\chi^2_{Int*Age}$=11.41, $p \leq 0.001$; χ^2_{Bg*Int}=9.18, $p \leq 0.01$), and *post-hoc* models with split factors were constructed to explore these interaction terms. Additional two-way, three-way, and four-way interaction terms were tested for significance using model testing, but none of these additional interaction terms significantly improved the fit of the model (χ^2<3.80, *ns* in all cases).

All subjects also participated in a similar study using interrupted PRESTO sentences (Experiment 2). The order of these experiments was randomized across subjects and a binary session order factor (*Ord*) was designated if data for Experiment 1 were collected in the first session (negative sign) or the second session (positive sign). This factor was added to the model and tested for significance but did not significantly improve the fit of the model (χ^2_{Ord}=2.4647; *ns*), indicating that previous experience with interrupted PRESTO sentences did not significantly improve keyword recognition for participants who had already completed Experiment 2.

Proportion of speech significantly interacted with background ($\beta_{Prop*Bg}$=0.10; z=6.62; $p \leq 0.001$). To explore this interaction, separate factors were generated for effects of speech proportion in quiet and with a competing talker (*Prop_Q* and *Prop_CT*); a model was generated in which these two factors replaced proportion of speech (*Prop*) and its interaction with background (*Prop*Bg*). Results of this split-factor model indicated that proportion was a stronger predictor of keyword recognition with the competing talker (β_{Prop_CT}=1.12; z=38.57; $p \leq 0.001$) than in quiet (β_{Prop_Q}=0.93; z=42.33; $p \leq 0.001$). The relative magnitudes of the standard estimates reflect the extent to which performance declined with decreasing speech proportion. As such, the source of the interaction in the base model was performance that declined more precipitously with decreasing speech proportion with the competing talker compared to quiet.

The interaction between proportion of speech and age group was also significant ($\beta_{Prop*Age}$=0.07; z=4.43; $p \leq 0.001$). Separate factors were generated for effects of speech proportion for younger and older adults (*Prop_Y* and *Prop_O*); a model was generated using these factors in place of the proportion factor and associated interaction (*Prop* and *Prop*Age*). Results of this split-factor model indicated that speech proportion was a stronger predictor of keyword recognition among older adults (β_{Prop_O}=1.06; z=42.79; p≤0.001) than younger adults (β_{Prop_Y}=0.92; z=39.13; p≤0.001). This suggests that performance declined more precipitously with decreasing speech proportion for older adults than younger adults.

Proportion of speech also significantly interacted with interruption type ($\beta_{Prop*Int}$=-0.07; z=-5.42; p≤0.001). Separate factors were generated for effects of speech proportion for silence-interrupted and envelope-filled sentences (*Prop_SInt* and *Prop_Env*); a model was generated using these factors in place of the proportion factor and associated interaction (*Prop* and *Prop*Int*). Results of this split-factor model indicated that proportion was a stronger predictor of keyword recognition for silence-interrupted sentences (β_{Prop_SInt}=1.06; z=42.14; p≤0.001) than envelope-filled sentences (β_{Prop_Env}=0.92; z=38.44; p≤0.001). This suggests that, with the addition of envelope-modulated noise, listeners were better able to use the available glimpses of speech to recognize the sentence. In the absence of these supportive envelope cues, keyword recognition depended more on the total amount of glimpsed speech information that was available to the listener.

Interruption type significantly interacted with age group ($\beta_{Int*Age}$=0.06; z=3.61; p≤0.001). Separate factors were generated for effects of interruption type among younger and older adults (*Int_Y* and *Int_O*); a model was generated in which these factors replaced interruption type and the associated interaction (*Int* and *Int*Age*). Results of this split-factor model indicated that the effect of interruption type was larger among older adults (β_{Int_O}=0.26; z=11.95; p≤0.001) than younger adults (β_{Int_Y}=0.16; z=7.59; $p \leq 0.001$). The relative magnitudes of the standard estimates reflect the amount of improvement associated with phonemic restoration. As such, the interaction between age and interruption type indicates a larger improvement associated with phonemic restoration among older adults compared to younger adults.

Finally, background significantly interacted with interruption type (β_{Bg*Int}=0.03; z=3.03; $p \leq 0.01$). Separate factors were generated for effects of background for silence-interrupted and envelope-filled sentences (*Bg_SInt* and *Bg_Env*); a model was generated in which these factors replaced background and the associated interaction (*Bg* and *Bg*Int*). Results of this split-factor model indicated that the competing talker had a more detrimental effect on keyword recognition for silence-interrupted (β_{Bg_SInt}=-0.73; z=-34.79; $p \leq 0.001$) than envelope-filled sentences (β_{Bg_Env}=-0.68; z=-33.58; $p \leq 0.001$). This suggests that the addition of envelope-modulated noise to silent intervals may have helped listeners segregate the glimpses of target speech from the competing talker.

II.C.2. Item-level factors

II.C.2.a. Lexical characteristics

Lexical characteristics included word frequency (*WF*), neighborhood density (*ND*), and biphone probability (*BP*). Initial modeling results indicated that all three lexical characteristics were correlated and the strongest predictor among the three factors was word frequency. To address collinearity, neighborhood density and biphone probability were residualized with respect to word frequency.

The resulting residualized factors were minimally correlated with each other and did not exceed the 0.3 correlation coefficient criterion. Subsequently, word frequency and residualized neighborhood density and biphone probability were normalized for mean of 0 and standard deviation of 1 and added to the GLMM. Model testing confirmed significant contributions of all 3 lexical characteristics (χ^2_{WF}=770.80, $p \leq 0.001$; χ^2_{ND}=201.51, $p \leq 0.001$; χ^2_{BP}=8.45, $p \leq 0.01$), as well as an interaction between word frequency and interruption type (χ^2_{Int*WF}=5.42, $p \leq 0.05$). Results indicated better recognition of commonly used keywords than less common words (β_{WF}=0.28; z=27.72; $p \leq 0.001$) and better recognition for keywords with more common phoneme sequences (β_{BP}=0.03; z=2.90; $p \leq 0.01$). Recognition was poorer for keywords from more dense neighborhoods (β_{ND}=-0.15; z=-14.00; $p \leq 0.001$). The significant interaction between word frequency and interruption type (β_{Int*WF}=-0.02; z=-2.32; $p \leq 0.05$) was explored with a *post-hoc* model with separate interruption type factors for more commonly used keywords (i.e., normalized $WF>0$; *Int_hiWF*) and less common keywords (i.e., normalized $WF<0$; *Int_loWF*). Results of the split-factor model indicated that phonemic restoration was greater for less common keywords (β_{Int_loWF}=0.25; z=18.64; $p \leq 0.001$) than for more common keywords (β_{Int_hiWF}=0.20; z=11.28; $p \leq 0.001$). Additional interactions, including interactions with age, were tested for significance but none of these interaction terms significantly improved the fit of the model (χ^2<1.67, *ns* in all cases).

II.C.2.b. Sentence length and keyword position

Sentence length (*nWords*) and keyword position (*Pos*) were normalized for mean of 0 and standard deviation of 1 and tested for significance in the GLMM. Model testing confirmed significant improvements in model fit with the addition of both factors (χ^2_{nWords}=156.19, $p \leq 0.001$; χ^2_{Pos}=270.35, $p \leq 0.001$). Keyword recognition declined for sentences with more keywords (β_{nWords}=-0.13; z=-12.45; $p \leq 0.001$), and recognition of later occurring keywords was better than for keywords occurring earlier in the sentence (β_{Pos}=0.17; z=16.38; $p \leq 0.001$). Three interaction terms also significantly improved model fit ($\chi^2_{Pos*Age}$=14.15, $p \leq 0.001$; $\chi^2_{nWords*Bg}$=63.54, $p \leq 0.001$; χ^2_{Pos*Bg}=102.64, $p \leq 0.001$) and these interactions were explored with *post-hoc* models with split factors.

The significant interaction between keyword position and age ($\beta_{Pos*Age}$=-0.04; z=-3.75; $p \leq 0.001$) was explored with separate keyword position factors for younger and older adults (*Pos_Y* and *Pos_O*). Results of the split-factor model indicated that keyword recognition improved more over the course of the sentence for younger adults (β_{Pos_Y}=0.19; z=13.64; $p \leq 0.001$) than for older adults (β_{Pos_O}=0.11; z=7.09; $p \leq 0.001$). This result was unexpected and may be a consequence of poorer performance for older adults in the more difficult conditions, such as with the competing talker or small proportions of speech. Sentences with no correct keywords were not uncommon for older adults in these conditions and these very poor scores effectively reduce the extent to which recognition can be predicted based on the position of the keyword in the sentence.

The significant interaction between sentence length and background ($\beta_{nWords*Bg}$=-0.09; z=-7.99; $p \leq 0.001$) was explored with separate sentence length factors for quiet and with a competing talker (*nWords_Q* and *nWords_CT*). Results of the split-factor model indicated more precipitous declines in recognition with increasing sentence length with a competing talker (β_{nWords_CT}=-0.22; z=-12.45; $p \leq 0.001$) than in quiet (β_{nWords_Q}=-0.05; z=-3.89; $p \leq 0.001$). This finding suggests that the additional demands associated with speech segregation compounded the difficulty associated with recognition of longer sentences.

The significant interaction between keyword position and background (β_{Pos*Bg}=0.11; z=10.12; $p \leq 0.001$) was explored with separate keyword position factors for quiet and with a competing talker (*Pos_Q* and *Pos_CT*). Results of the split-factor model indicated greater improvements in recognition over the course of the sentence with a competing talker (β_{Pos_CT}=0.28; z=17.10; $p \leq 0.001$) than in quiet (β_{Pos_Q}=0.07; z=5.28; $p \leq 0.001$). This finding is consistent with the hypothesis that speech segregation develops slowly over time, resulting in better recognition of keywords occurring later in the sentence. However, the three-way interaction between keyword position, background, and age did not significantly improve the model (χ^2=0.45, *ns*), suggesting that the time course of speech segregation was similar for younger and older adults.

II.C.2.c. Talker sex

The fixed effect of talker sex did not significantly improve model fit (χ^2_{Sex}=0.02, *ns*), but significant improvements in model fit were observed with interactions between talker sex and age ($\chi^2_{Sex*Age}$=51.93, $p\leq0.001$), talker sex and background (χ^2_{Sex*Bg}=11.17, $p\leq0.001$), and a three-way interaction between talker sex, background, and age ($\chi^2_{Sex*Bg*Age}$=16.01, $p\leq0.001$). The interaction between talker sex and listener age ($\beta_{Sex*Age}$=-0.08; z=-7.95; $p\leq0.001$) was explored with separate talker sex factors for younger and older adults (*Sex_Y* and *Sex_O*). Results of this split-factor model revealed β_{Sex} coefficients with opposite signs, indicating opposite effects of talker sex for younger and older adults. For younger adults, keyword recognition was better when the talker was female (β_{Sex_Y}=0.07; z=5.29; $p\leq0.001$), whereas for older adults, keyword recognition was better when the talker was male (β_{Sex_O}=-0.07; z=-4.70; $p\leq0.001$). These equivalent magnitude effects with opposite signs likely cancelled out the fixed effect of talker sex in the original model. The interaction between talker sex and background (β_{Sex*Bg}=0.03; z=3.03; $p\leq0.01$) was explored with separate talker sex factors in quiet and with a competing talker (*Sex_Q* and *Sex_CT*). Results of this split-factor model revealed similar opposite-sign effects; recognition of male talkers was better than female talkers in quiet (β_{Sex_Q}=-0.03; z=-2.54; $p\leq0.05$), whereas recognition of female talkers was better than male talkers with a competing talker (β_{Sex_CT}=0.04; z=2.18; $p\leq0.05$). The combined effects of these two interactions produced a significant three-way interaction ($\beta_{Sex*Bg*Age}$=-0.04; z=-4.00; $p\leq0.001$), which was explored by splitting the talker sex factor twice, first by background and then by age group, producing 4 factors (*Sex_Q_Y, Sex_Q_O, Sex_CT_Y,* and *Sex_CT_O*). These factors were added to a *post-hoc* model in place of talker sex and both 2-way interactions. Results revealed that, in the quiet background, talker sex did not affect recognition for younger adults ($\beta_{Sex\_Q\_Y}$=0.01; z=0.53; *ns*), but older adults had greater difficulty with female talkers compared to male talkers ($\beta_{Sex\_Q\_O}$=-0.07; z=-3.91; $p \leq 0.001$). With a competing talker, younger adults performed better with a female talker compared to a male talker ($\beta_{Sex\_CT\_Y}$=0.15; z=7.02; $p \leq 0.001$), whereas older adults performed poorer with a female talker ($\beta_{Sex\_CT\_O}$=-0.09; z=-3.84; $p \leq 0.001$).

II.C.3. Subject-level factors

II.C.3.a. Education level

Younger and older subjects had similar years of education (education level). Mean education level was 16.8 years (*SD*=2.3) for younger subjects and 16.4 years (*SD*=2.4) for older subjects. Independent samples *t* test did not reveal significant differences in education level between younger and older subjects ($t(37.88)=0.54$, *ns*). As such, education level (*Edu*) was normalized across all subjects for a mean of 0 and a standard deviation of 1 and added to the GLMM. Model testing confirmed a significant improvement in model fit related to education level (χ^2=5.16; $p \leq 0.05$), such that keyword recognition increased significantly with increasing education level (β_{Edu}=0.10; z=2.35; $p \leq 0.05$).

II.C.3.b. Hearing sensitivity

Effects of hearing sensitivity were evaluated based on 4 PTA measures; narrow PTA (NPTA: 0.5, 1.0, 2.0, 4.0 kHz), broad PTA (BPTA: 0.5, 1.0, 2.0, 3.0, 4.0, 6.0, 8.0 kHz), low-frequency PTA (LFPTA: 0.25, 0.5, 1.0 kHz), and high-frequency PTA (HFPTA: 2.0, 3.0, 4.0, 6.0, 8.0 kHz). Independent samples *t* test were performed on PTA measures for younger and older adults using Bonferroni correction for multiple comparisons. Results indicated that NPTA for younger adults (*M*=2.1 dB HL, *SD*=4.0) was significantly lower than for older adults (*M*=10.2 dB HL, *SD*=3.6; *t*(37.53)=6.71, $p \leq 0.001$), BPTA for younger adults (*M*=3.3 dB HL, *SD*=3.7) was significantly lower than older adults (*M*=15.2 dB HL, *SD*=4.9; *t*(35.06)=8.68, $p \leq 0.001$), LFPTA for younger adults (*M*=3.1 dB HL, *SD*=2.9) was significantly lower than for older adults (*M*=7.8 dB HL, *SD*=4.3; *t*(33.18)=4.11, $p \leq 0.001$), and HFPTA for younger adults (*M*=3.5 dB HL, *SD*=4.5) was significantly lower than for older adults (*M*=18.1 dB HL, *SD*=6.7; *t*(33.06)=8.10, $p \leq 0.001$). These group differences resulted in collinearity of effects when PTA and age group factors were included in the GLMM. To determine if differences in hearing sensitivity were driving the observed effects of age, models were generated in which the age factor was replaced with each of the 4 PTA measures (normalized for *M*=0, *SD*=1). Comparisons of β values across these models indicated that age was a better predictor of keyword recognition than any measure of PTA (β_{Age}=-0.29; β_{NPTA}=-0.24; β_{BPTA}=-0.26; β_{LFPTA}=-0.23; β_{HFPTA}=-0.26). Next, each PTA measure was residualized for age group differences and the residualized PTA factors were added to the original GLMM and tested for significance using model testing. None of the residualized factors significantly improved model fit (χ^2_{NPTA}=1.12; χ^2_{BPTA}=1.58; χ^2_{LFPTA}=1.68; χ^2_{HFPTA}=1.83; *ns*), indicating that differences in hearing sensitivity among younger and older adults were not predictive of keyword recognition.

II.C.3.c. Cognitive measures

Cognitive test results for younger and older adults were transformed into *z*-scores and are displayed in Figure 6. Independent samples *t* test were performed on each cognitive measure for younger and older adults using Bonferroni correction for multiple comparisons. Results indicated younger adults outperformed older adults for all cognitive measures (Connections: *t*(37.10)=5.21, *p≤0.001*; Peg Board: *t*(37.82)=4.91, *p≤0.001*; Reading Span: *t*(35.26)=3.26, *p≤0.05*; Stroop: *t*(32.47)=3.60, *p≤0.01*; TRT: *t*(26.88)=4.41, *p≤0.001*).

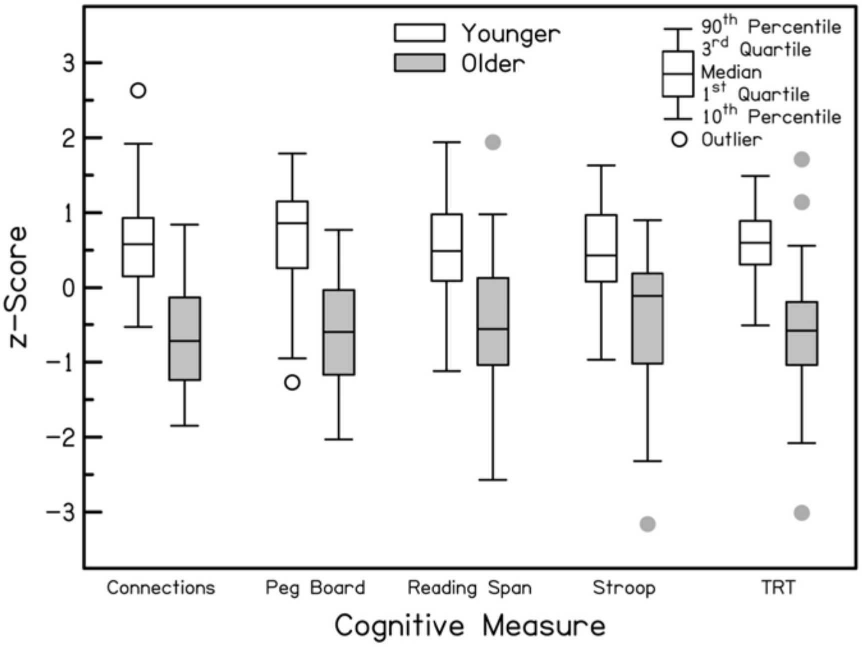

Figure 6: Cognitive test results, transformed to *z*-scores, for younger (white) and older (gray) participants. Independent samples *t* tests with Bonferroni correction confirm age group differences are significant for all tests (*p*<0.05).

Group differences resulted in collinearity of effects when cognitive factors and age group were included in the GLMM. To address this issue, each cognitive measure was residualized for age group differences and these residualized factors were added to the GLMM and tested for significance using model testing. Initial modeling results revealed significant collinearity between TRT and education level. To address this issue, TRT was further residualized for differences in education level. Final modeling results indicated significant improvements in model fit with the addition of Connections (χ^2=7.15; $p \leq 0.01$), TRT (χ^2=6.31; $p \leq 0.05$) and an interaction between TRT and interruption type (χ^2=14.84; $p \leq 0.001$). Results indicated that keyword recognition improved with faster speed of processing ($\beta_{Connections}$=0.13; z=2.80; $p \leq 0.01$) and better linguistic closure (β_{TRT}=0.13; z=2.60; $p \leq 0.01$). The significant interaction between TRT and interruption type ($\beta_{TRT*Int}$=-0.06; z=-4.19; $p \leq 0.001$) was explored with a *post-hoc* model with split factors for TRT with silence-interrupted (*TRT_SInt*) and envelope-filled sentences (*TRT_Env*); results indicated that TRT significantly predicted keyword recognition for silence-interrupted sentences (β_{TRT_SInt}=0.19; z=3.40; $p \leq 0.001$) but not for envelope-filled sentences (β_{TRT_Env}=0.07; z=1.44; *ns*). The same interaction term was explored further by splitting the interruption type factor for subjects with high TRT scores (i.e., *TRT*>0; *Int_hiTRT*) and low TRT scores (i.e., *TRT*<0; *Int_loTRT*); results of this split-factor model indicated that subjects with lower TRT scores demonstrated larger improvements associated with phonemic restoration (β_{Int_loTRT}=0.27; z=12.40; $p \leq 0.001$), than subjects with higher TRT scores (β_{Int_hiTRT}=0.21; z=9.53; $p \leq 0.001$). This result is not surprising, in light of the finding that better TRT scores predicted better performance in the baseline (silence-interrupted) condition. Other cognitive factors and interactions between cognitive and design-level factors such as age and background were explored, but model testing revealed that these additional factors did not improve the fit of the model (χ^2<1.40; *ns* in all cases).

II.D. Discussion

This study investigated the relative contributions of glimpsing, phonemic restoration, and speech segregation to recognition of interrupted speech in younger and older adults with normal hearing. As expected, older adults performed poorer than younger adults across all speech conditions. Similar age-related declines in recognition of interrupted speech, even in the absence of peripheral hearing loss, have been observed in several other studies (Gordon-Salant & Fitzgibbons, 1993; Krull et al., 2013; Shafiro et al., 2015). This study included only older adults with normal hearing to minimize effects of reduced audibility on speech recognition. Although older adults had slightly elevated mean thresholds compared to the younger adults, variance in hearing sensitivity measured across several different frequency ranges was not found to significantly predict keyword recognition among younger or older adults. Rather, several other factors not related to pure-tone thresholds were found to contribute to keyword recognition and to the effects of age. These findings shed light on the extent to which age-related declines in speech recognition can be explained by declines in glimpsing, phonemic restoration, and/or speech segregation.

II.D.1. Glimpse duration and proportion of speech

The ability to use glimpsed speech information to recognize the sentence was operationalized here as speech proportion. Sentences were processed such that equal duration segments were removed, and the duration of glimpses increased with the proportion of speech. Increasing speech proportion also affected the interruption rate, resulting in fewer overall interruptions, which may have contributed to the improvement in speech recognition with increasing proportion of speech. As such, the increase in speech recognition with increasing proportion of speech reflects the extent to which longer glimpses with fewer interruptions facilitated better keyword recognition. Previous studies in which interruption rate, glimpse duration, and speech proportion were varied orthogonally indicated that overall speech proportion is the best predictor of recognition of interrupted speech (Gilbert et al., 2007; Wang & Humes, 2010). Although these factors covaried in the present investigation, proportion of speech was observed to be the strongest predictor of performance. Here, proportion was also observed to interact with several other factors. The pattern of these interactions was consistent; that is, any additional factor that decreased performance (competing talker, older age, silent interruptions) resulted in a stronger effect of speech proportion. Thus, factors that increased the difficulty of the task also increased the contributions of glimpse duration and the available speech information to overall recognition of the sentence. The competing talker likely resulted in some amount of energetic masking that reduced the “effective” proportion of speech. That, coupled with the added requirement of segregating glimpses from the competing talker, resulted in greater reliance on available speech information. The addition of envelope-modulated noise allowed listeners to more effectively use the available glimpses of speech and connect those glimpses over time. The improvements in speech recognition that occurred when interruptions were filled with envelope-modulated noise (phonemic restoration) effectively minimized the importance of the available glimpses, allowing both groups to recognize more keywords with shorter glimpses of speech.

One hypothesis with respect to the effects of age on speech recognition predicts age-related declines in the ability to effectively use glimpsed speech information. Some support for this hypothesis can be drawn from the negative effect of age on overall performance, as has been shown previously (Gordon-Salant & Fitzgibbons, 1993; Krull et al., 2013; Shafiro et al., 2015). Further support can be drawn from the observed interaction between age and speech proportion. The source of this interaction was a larger effect of speech proportion for older than younger adults, indicating an age-related decline in glimpsing. Older adults relied more heavily on the available speech information, whereas younger adults were able to make more efficient use of that available information to recognize more keywords. As the speech proportion increased, keyword recognition for younger and older adults converged, indicating that the negative consequence of age depended critically on the duration of glimpses and overall proportion of speech. This is reflected in the data, where the difference in scores between younger and older adults is greatest for smaller speech proportions, and converges for the largest speech proportion. These results indicate an age-related decline in glimpsing, such that older adults were less adept at recognizing short glimpses of speech and connecting them together across time into a single coherent message.

II.D.2. Perceptual “cost” of speech segregation

Speech recognition was substantially poorer with the competing talker compared to quiet. This decline in recognition reflects the combined effects of energetic masking and the additional perceptual demands associated with segregation of speech glimpses from a competing talker. Although the effects of energetic masking cannot be ignored, the masking effects of a single competing talker are thought to predominantly reflect the costs associated with speech segregation and selection of the appropriate target talker (i.e., “informational masking;” Brungart, 2001; Kidd et al., 2005). In addition, the two item-level factors that interacted with background both related to the time course of the sentence, keyword position and sentence length. These results revealed that with the competing talker, longer sentences were generally more difficult but listeners were more likely to correctly identify keywords as the sentence progressed. Whereas effects of keyword position and sentence length were also significant in the quiet background, the magnitude of their effects was substantially smaller than with a competing talker. Effects of the competing talker associated with energetic masking would be expected to be equivalent for sentences of any length and remain essentially stable throughout the sentence. Thus, the observed interactions between keyword position, sentence length, and background support the interpretation that the effect of the competing talker was an increase in task demands associated with speech segregation.

Our hypothesis predicted that age-related declines in speech recognition would be explained by difficulty segregating speech from the competing talker.

This prediction was based on results of several studies that demonstrated poorer speech recognition by older adults compared to younger adults in competing talker backgrounds (e.g., Helfer & Freyman, 2008; Lee & Humes, 2012) and a prolonged time course of speech segregation for older adults (Ben-David et al., 2012; Ezzatian et al., 2015). As such, the decline in performance with the addition of a competing talker was expected to be greater for older adults than younger adults and the improvement in performance over the course of the sentence with a competing talker was expected to be greater for older adults than younger adults. Neither of these interactions was supported by the data. Recognition of silence-interrupted speech is a difficult task, particularly for older adults, and the addition of a competing talker may have resulted in floor performance for some older subjects. As a result, effects of a competing talker among older adults may have been underestimated in this experiment. The effect of keyword position observed here may reflect factors other than speech segregation. For example, the benefits associated with context generally improve over the course of a sentence (Cole & Perfetti, 1980; Kidd & Humes, 2012) and these context-dependent effects may have been more beneficial in the more difficult competing talker background (Dirks, Bell, Rossman, & Kincaid, 1986).

Poorer performance for longer sentences with a competing talker is suggestive of effects related to working memory capacity, but this interpretation was not supported by subject-level differences in performance on the Reading Span. Sentences with more keywords were expected to place greater demands on working memory, which may have contributed to the overall negative effect of sentence length on performance. Subjects with higher working memory capacity were expected to leverage their strong cognitive abilities to facilitate better speech recognition, particularly in the more difficult competing talker background. However, performance on the Reading Span did not predict keyword recognition or influence the effect of sentence length on performance. One possibility is that the visual measure of working memory capacity was not sensitive to the effects of working memory on the auditory speech recognition task. Stronger associations may have been observed with an auditory measure of working memory capacity (e.g., Gordon-Salant & Cole, 2016).

Speech segregation was more difficult for silence-interrupted sentences than envelope-filled sentences. There are at least two possible interpretations for this finding. First, silent intervals may serve as openings for listeners to confuse competing talker segments for glimpses of target speech (Gaudrain & Carlyon, 2013; Gnansia et al., 2010). Providing a continuous envelope cue by filling silent intervals with envelope-modulated noise may have helped listeners track the envelope of the target and avoid confusing the competing talker with the target speech. Alternatively, or perhaps additionally, the envelope-modulated noise may have served as an energetic masker of the competing talker during segments between speech glimpses. Previous studies involving energetic masking of a competing talker have shown a reduction in the “informational masking” component of competing speech when it is masked by noise (Agus, Akeroyd, Gatehouse, & Warden, 2009; Kidd et al., 2005). These two interpretations are not mutually exclusive and both may have contributed to the observed difference in speech segregation for silence-interrupted and envelope-filled sentences.

II.D.3. Effects of age on phonemic restoration

Older adults demonstrated a larger benefit than younger adults when silent interruptions were filled with envelope-modulated noise. A similar finding was observed in a subset of conditions in recent work by Saija and colleagues (2014). In their study, enhanced phonemic restoration was observed for older adults with normal rate and time-expanded speech. They suggested that older adults may be able to compensate for age-related declines in speech recognition by relying on their linguistic skills and vocabulary and/or by exerting additional cognitive effort to fill in missing information. In this study, the contributions of lexical factors of keywords and cognitive abilities were assessed with item-level and subject-level factors to test these hypotheses.

Analyses of lexical factors of keywords and subject-level differences in education did not support the hypothesis that language abilities contribute to enhanced phonemic restoration. Item-level variation in word frequency, biphone probability, and neighborhood density were each predictive of keyword recognition, but only word frequency affected phonemic restoration. This effect was relatively small and no associations were observed between any of the lexical factors and age. Whereas vocabulary or language abilities of subjects were not directly assessed, subjects with more years of education likely had larger vocabularies and better language abilities than subjects with fewer years of education (Verhaeghen, 2003). Although performance improved with increasing years of education, younger and older subjects did not differ in education level, and this factor did not influence the improvement associated with phonemic restoration. As such, these results provide little support for the hypothesis that strong language abilities of older adults contribute to enhanced phonemic restoration.

The contributions of cognitive abilities to phonemic restoration were evaluated with a battery of cognitive tests, including the TRT, which assessed the ability to make use of partial linguistic information in the visual modality. Better performance on the TRT predicted better recognition of silence-interrupted sentences, in agreement with results from Krull and colleagues (2013). The task of reading a visually obscured sentence is intuitively similar to recognition of interrupted speech and the association between these two measures suggests a common amodal ability to make use of partial linguistic information (Krull et al., 2013). However, older subjects performed poorer on the TRT overall and subjects with better TRT scores demonstrated *less* phonemic restoration than those with poorer TRT scores. These findings remained consistent even after TRT scores were residualized for differences in age and education level. This suggests that greater cognitive effort needed to fill in missing information is not likely to explain increased phonemic restoration among older adults. If anything, efforts to fill in missing information are likely to improve recognition of silence-interrupted speech, thereby reducing the amount of improvement that is observed with phonemic restoration. Other cognitive measures were not found to predict keyword recognition, with the exception of the Connections Test. These results indicated that faster processing speed contributed to better keyword recognition, but did not influence the degree of improvement with phonemic restoration.

The observed improvement in phonemic restoration with advancing age may be a consequence of age-related declines in temporal resolution and glimpsing. In this study, older adults were particularly poor at recognizing short glimpses of speech and connecting them across time. Glimpsing, in this context, may involve a form of template matching in which the internal representation of the glimpsed signal is compared to several possible words that may match the stimulus (i.e., Srinivasan & Wang, 2005). As such, a critical ability for recognition of glimpsed speech may be maintaining an accurate internal representation of the glimpses and their relative position across time. To the extent that the listener retains the precise temporal relationships between glimpses, the ability to match their internal representation of the signal to the appropriate speech template would be optimized. Age-related declines in temporal resolution are known to compromise the ability to maintain temporal relationships by older adults (Fitzgibbons & Gordon-Salant, 1995; 1996; 2001) and particular difficulty has been noted among older adults for recognition of silent temporal intervals (Fitzgibbons & Gordon-Salant, 1987; 1994; Gordon-Salant, Yeni-Komshian, Fitzgibbons, & Barrett, 2006). As such, age-related declines in temporal resolution may be particularly disruptive to recognition of silence-interrupted speech. Inserting noise into silent intervals may provide an auditory scaffold that helps older adults maintain an accurate temporal representation of the glimpses across the length of the sentence. Therefore, enhanced phonemic restoration observed for older adults may reflect the combined benefits of phonemic restoration and improved temporal representation of the glimpses relative to silence-interrupted speech.

In Experiment 1, intervening noise contained random fluctuations in amplitude and was also modulated by the temporal envelope of the missing speech segments. The intrinsic fluctuations in amplitude likely produced some degree of modulation masking and may have interfered with recognition of important envelope cues in the glimpses (Stone et al., 2012). However, the addition of noise to silent intervals improved sentence recognition, suggesting that any modulation masking effects that may have occurred were small relative to the improvement in recognition associated with phonemic restoration. Modulation of the noise by the broadband envelope of the missing speech segments has been shown to enhance phonemic restoration relative to steady-state noise, which contains no beneficial envelope cues (Bashford et al., 1996; Shinn-Cunningham & Wang, 2008). These envelope cues may have further improved the ability to connect glimpses of speech across time by providing a predictable pattern of modulations connecting one glimpse to the next. These envelope cues may have also enhanced the process of template matching, as the entire broadband temporal envelope of the sentence was available for comparison to possible words and phrases in the listener's lexicon. Results of lexical factors provide some support for this interpretation. Word frequency was a better predictor of recognition of silence-interrupted sentences than envelope-filled sentences, indicating that envelope cues facilitated recognition of less common keywords. Thus, the additional envelope cues in filled sentences likely contributed to the restoration effects observed in Experiment 1.

II.D.4. Effects of age and talker sex

Older adults performed poorer on sentences with a female talker than a male talker in quiet and with a competing talker. The PRESTO sentence material includes a diverse sample of male and female talkers from different dialect regions and it is therefore unlikely that these effects were related to differences in intelligibility for specific male and female talkers in the corpus. Older adults in this study had hearing thresholds ≤ 30 dB HL through 6.0 kHz and the effect of talker sex was not predicted by variance in any PTA measure. Thus, these results suggest that age-related changes in the auditory system other than reduced speech audibility may disrupt the ability to understand female talkers more so than male talkers. The precise source of this decline remains unclear. Poorer coding of faster rate periodicity cues in older adults (Clinard & Tremblay, 2013; Snyder & Alain, 2005) may have a more detrimental effect on the neural representation of female voices than male voices. Poorer spectral representation of F0 in the harmonic structure of speech may also result in poorer speech recognition and fewer opportunities for release from masking release with a competing talker (Deroche, Culling, Chatterjee, & Limb, 2014; Deroche, Culling, & Chatterjee, 2014). The contributions of periodicity cues to speech recognition, as well as the effect of periodicity cues on recognition of male and female talkers, was further explored in Experiment 2.

Younger adults displayed similar recognition of male and female talkers in quiet, but demonstrated an advantage for recognition of female talkers with competing speech. The source of this talker sex effect is also unclear. Owing to the design of the study, female target sentences were always presented with a male competing talker, and *vice-versa*. As such, it is not possible to distinguish better recognition of female target talkers from less effective masking by male competing talkers. Future studies should employ additional conditions with same sex maskers to further explore talker sex differences.

II.E. Conclusions

This study used an interrupted speech paradigm to determine the extent to which age-related declines in speech recognition could be explained by difficulty with glimpsing, phonemic restoration, and/or speech segregation. Poorer speech recognition was observed for older adults compared to younger adults and this age-related decline could not be explained by variance in hearing sensitivity. Poorer speech recognition among older adults was partially explained by declines in glimpsing, particularly for silence-interrupted speech. Phonemic restoration, measured as the improvement in speech recognition when silent intervals were filled with noise, was greater in older adults than younger adults. The additional envelope cues available in envelope-filled sentences may have provided an auditory scaffold that helped facilitate glimpsing in older adults. The addition of an opposite-sex competing talker resulted in poorer performance for both groups, particularly for silence-interrupted sentences. However, no age-related differences were observed in the extent of this decline. Larger effects of age were observed for sentences with female target talkers than male talkers, indicating that older adults may be poorer at recognizing speech with faster rate periodicity cues and broader harmonic structure. Subject-level differences in processing speed and linguistic closure were predictive of better keyword recognition, indicating that recognition of interrupted speech requires quick and efficient use of partial linguistic information.

III. Experiment 2: Contributions of Envelope and Periodicity Cues to Recognition of Speech Glimpses for Younger and Older Adults

III.A. Introduction

Speech contains distinct information across a range of temporal fluctuation rates, which may facilitate recognition in realistic listening environments. The framework developed by Rosen (1992) distinguishes envelope cues (2-50 Hz), which code syllabic and segmental rates, from periodicity cues (50-500 Hz), which code F0 and intonation. The contributions of envelope cues to speech recognition are considerable, as evidenced by the high intelligibility of vocoded speech (Shannon et al., 1995). However, envelope cues are generally similar across talkers, making them highly susceptible to modulation masking effects in realistic listening environments (Kwon & Turner, 2001). Periodicity cues generally differ between talkers, making them particularly useful for speech segregation with a competing talker (Brokx & Nooteboom, 1982; Stickney et al., 2007). Periodicity cues also convey intonation, which varies dynamically over the course of the sentence and may provide a predictable pitch pattern that can be used as a basis for glimpsing (Gaudrain et al., 2007; Woods & McDermott, 2015). These findings suggest that in realistic listening environments, listeners may rely on periodicity cues for glimpsing segments of speech while using envelope cues as the basis for speech recognition. The purpose of Experiment 2 was to determine the extent to which younger and older adults can use envelope and periodicity cues for recognition of glimpsed speech in quiet and with a competing talker.

Converging sources of evidence suggest that aging may negatively affect the ability to use periodicity cues to facilitate speech recognition. Age-related declines in the neural representation of periodicity cues in the brainstem have been demonstrated using electrophysiologic methods (Clinard & Tremblay, 2013; Snyder & Alain, 2005). Poorer coding of periodicity cues in older adults likely contributes to poorer performance on several speech-based tasks, such as concurrent vowel identification (Arehart et al., 2011; Chintanpalli et al., 2016; Vongpaisal & Pichora-Fuller, 2007). Pichora-Fuller and colleagues investigated the effects of a simulated decline in periodicity coding on recognition of speech in noise using sentences that were temporally “jittered” at frequencies below 1.2 kHz (Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007). Word recognition in noise was reduced in younger adults with normal hearing when low-frequency speech information was temporally jittered. Furthermore, performance closely matched that of older adults with normal hearing listening to stimuli without a temporal jitter. These results suggest that age-related declines in periodicity coding may contribute to the difficulty older adults experience in realistic listening environments.

Periodicity cues are thought to be particularly important for speech segregation and thus age-related declines in periodicity coding may explain the difficulty older adults experience on speech recognition tasks with competing talkers (e.g., Helfer & Freyman, 2008; Lee and Humes, 2012; Rajan & Cainer, 2008). In Experiment 1, performance declined equivalently for younger and older adults with addition of a competing talker, which may have been due to the high degree of difficulty associated with segregating silence-interrupted and noise-filled speech from a competing talker. Filling the silent intervals of interrupted speech with periodicity cues may help facilitate speech segregation, resulting in greater benefit of periodicity cues than envelope cues with a competing talker. However, age-related declines in periodicity coding may limit the benefit older adults can gain from using periodicity cues for speech segregation. Periodicity is also represented in the harmonic structure of voiced speech segments, which allows additional opportunities for “spectral glimpsing” between resolved harmonics of a competing talker (Deroche et al., 2014). Age-related declines in the use of these spectral cues may also limit the benefits of periodicity in older adults (Souza, Arehart, Miller, & Muralimanohar, 2011).

Envelope and periodicity cues may provide separate contributions to speech recognition and their combined effects may be redundant, additive, or synergistic. Several studies have demonstrated that the addition of F0-based periodicity information to vocoded speech enhances speech recognition (Başkent & Chatterjee, 2010; Oh et al., 2016; Zhang, Dorman, & Spahr, 2010). With a competing talker, the benefit of periodicity cues increases with the number of envelope channels in the vocoder, suggesting a synergistic effect between envelope and periodicity cues (Stickney et al., 2007). However, studies using more natural speech stimuli have not consistently shown a decline in glimpsing or speech segregation when periodicity cues are disrupted (Clarke et al., 2014; Freyman, Griffin, & Oxenham, 2012; Oxenham & Simonson, 2009). As such, it remains unclear the extent to which recognition of interrupted speech may improve when both envelope and periodicity cues are available. Results of Experiment 1 indicated that providing continuous envelope cues over the course of a sentence improved recognition of interrupted speech. Experiment 2 investigated the extent to which a continuous source of periodicity cues would result in a similar benefit for recognition and whether combining envelope and periodicity cues would provide any additional benefit compared to either cue alone.

The goals of Experiment 2 were to determine the extent to which aging affects the ability to use periodicity cues for speech recognition in quiet and with a competing talker and whether envelope and periodicity cues provide redundant, additive, or synergistic benefits when both cues are available. Younger and older adults with normal hearing listened to sentences interrupted with steady-state pulse trains that carried periodicity cues without envelope modulations, or envelope-modulated pulse trains that carried both envelope and periodicity cues. The proportion of the sentence remaining after interruption was manipulated based on the same speech proportion values used in Experiment 1, such that the contributions of envelope and periodicity cues to glimpsing could be evaluated across both experiments. Sentences were presented in quiet and with a competing talker to determine the extent to which envelope and periodicity cues contributed to speech segregation. Younger adults were expected to benefit more from periodicity cues with a competing talker than in quiet and older adults were expected to demonstrate less benefit overall due to age-related declines in periodicity coding. The addition of envelope and periodicity cues was expected to provide an additive benefit, relative to sentences from Experiments 1 and 2 containing each cue in isolation. As in Experiment 1, interactions and additional contributions from item-level factors (lexical characteristics, sentence length and keyword position, and talker sex) and subject-level factors (education level, hearing sensitivity, and cognitive abilities) were explored within a GLMM to explore potential interactions with periodicity cues and/or the combined benefits of envelope and periodicity cues. These results revealed the extent to which aging affects the relative and combined benefits of envelope and periodicity cues for speech recognition in realistic listening environments.

III.B. Methods

III.B.1. Subjects

The same 40 subjects from Experiment 1 participated in Experiment 2 so that data could be pooled across the two experiments to determine the relative and potentially additive contributions of envelope and periodicity cues to keyword recognition.

III.B.2. Stimuli and apparatus

Two additional sets of interrupted sentences were generated with identical speech proportions as described in Experiment 1, except that silent intervals were filled with pulse trains rather than noise. Pulse trains consisted of a series of periodic pitch pulses carrying the time-varying F0 from the original sentence and were generated using the Praat software package (Boersma & Weenick, 2014). Pitch contours were extracted from each sentence with a sampling rate of

100 Hz. This process occasionally resulted in aberrant high-frequency values being interpreted as part of the pitch contour and these aberrant points were removed by hand and replaced with actual pitch values as needed. Pitch contours were interpolated through unvoiced segments and periods of silence to generate continuous periodic pulse trains with a flat spectrum that followed the pitch contour of the original sentence. These pulse trains were subsequently filtered by the LTASS described in Experiment 1 and then amplified or attenuated as needed to match the RMS of the original sentence. These continuous pulse trains were then gated and combined with silence-interrupted sentences to generate periodicity-filled sentences. A second set of pulse trains was modulated by the envelope of the original sentence (as described for noise in Experiment 1) and then gated and combined with interrupted speech to generate envelope-plus-periodicity-filled sentences. The same competing talker sentences used in Experiment 1 were mixed with target sentences at the same SNR (+3 dB) for the competing talker conditions. Calibration procedures and presentation apparatus was identical to Experiment 1.

Example waveforms and spectrograms for the four types of interrupted sentences in Experiments 1 and 2 are displayed in Figure 7; the left two panels illustrate silence-interrupted and envelope-filled sentences from Experiment 1 and the right two panels illustrate periodicity-filled and envelope-plus-periodicity-filled sentences from Experiment 2. These 4 sentence types allowed for evaluation of the separate contributions of envelope and periodicity cues, as well as their combined effect. Sentence types displayed in the bottom panels of

Figure 7 (envelope filled and envelope-plus-periodicity filled) contained envelope cues, whereas sentences types displayed in the right panels of Figure 7 (periodicity filled and envelope-plus-periodicity filled) contained periodicity cues. This allowed for envelope and periodicity cues to be evaluated in a GLMM across the two experiments based on 2 binomial factors with a 2x2 design.

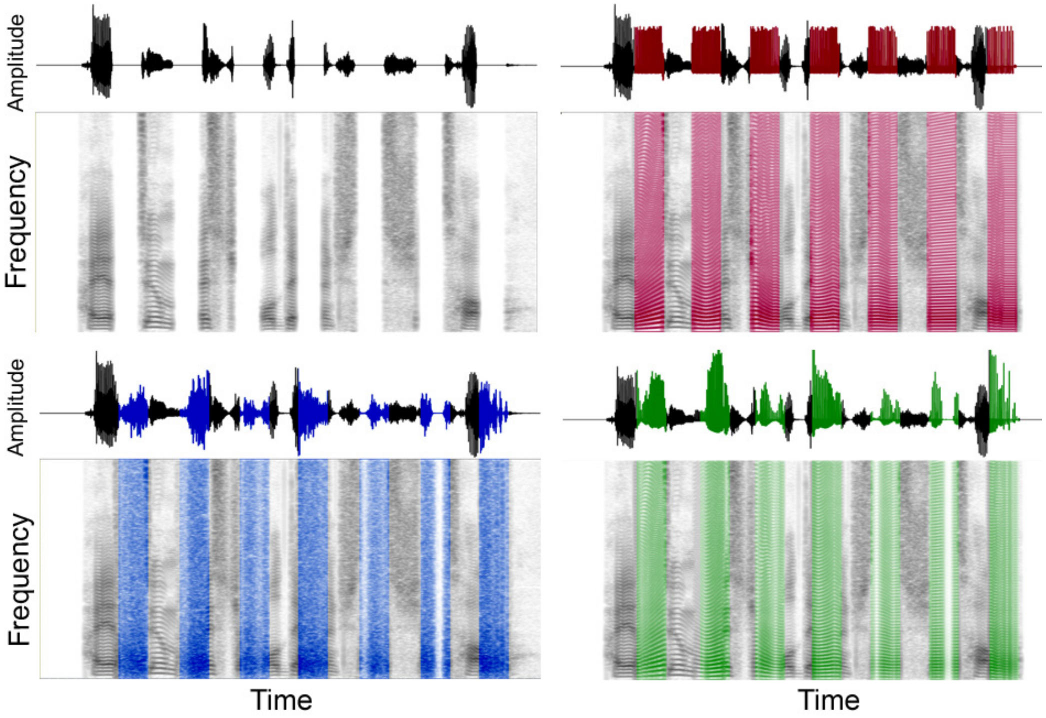

Figure 7: Example waveforms and spectrograms for the four types of interrupted sentences. Top left panel: Silence-interrupted sentence contains neither envelope nor periodicity cues between speech glimpses. Bottom left panel: Envelope-filled sentence contains envelope cues, but not periodicity cues. Top right panel: Periodicity-filled sentence contains periodicity cues, but not envelope cues. Bottom right panel: Envelope-plus-periodicity-filled sentence contains both envelope and periodicity cues.

III.B.3. Procedures

Procedures for speech recognition testing were identical to Experiment 1. Testing was blocked by background (quiet/competing talker) and counterbalanced across subjects. Data collection in each background consisted of one list at each speech proportion for periodicity-filled and envelope-plus-periodicity-filled sentences in a random order. A break was offered between quiet and competing talker test blocks. Speech testing was completed in a single session lasting about two hours. The order of participation in Experiments 1 and 2 was randomized across subjects to minimize any systematic effects of practice or familiarity with the sentences.

III.B.4. Statistical approach

Speech data collected in Experiment 2 were added to data from Experiment 1 and analyzed using a logistic regression GLMM. Two binary factors described the presence of envelope cues (*Env*) and periodicity cues (*Prd*) during sentence interruptions in a 2x2 design. Sentences from Experiment 2 were coded with a positive sign for *Prd* to indicate the presence of periodicity cues. Sentences filled with envelope-modulated signals (noise and pulse trains) were coded with a positive sign for *Env*, to indicate the presence of envelope cues. As such, envelope-plus-periodicity-filled sentences were coded with positive sign for both *Env* and *Prd*, while silence-interrupted sentences from Experiment 1 were coded with a negative sign for *Env* and *Prd*. Thus, across the 4 sentence types in Experiments 1 and 2, all combinations of envelope and periodicity cues were evaluated. The effect of envelope cues (*Env*) reflects performance for envelope-filled and envelope-plus-periodicity-filled sentences, as compared to silence-interrupted and periodicity-filled sentences. Likewise, the effect of periodicity cues (*Prd*) reflects performance for periodicity-filled and envelope-plus-periodicity-filled sentences, as compared to silence-interrupted and envelope-filled sentences.

All subjects participated in two experiments containing interrupted PRESTO sentences, and as a result, performance may have improved in the second testing session due to practice or familiarity effects. As such, a binary session order factor (*Ord*) was designated if data were collected in the first session (negative sign) or the second session (positive sign) and this factor was added to the model and tested for significance along with other design-level factors. Model testing revealed that session order significantly improved the fit of the model ($\chi^2_{Ord}=13.00$, $p\leq 0.001$); modeling results indicated a modest, but significant, improvement in keyword recognition on the second test session compared to the first ($\beta_{Ord}=0.05$, $z=3.92$, $p<0.001$). Interactions were explored between session order and proportion, background, envelope cues, periodicity cues, and age, but none of these interactions significantly improve the fit of the model ($\chi^2_{Ord}<1.73$, *ns* in all cases). This session order effect on keyword recognition was minor, corresponding to an average improvement of 3% across subjects for keyword recognition on the second session compared to the first session. As the size of the effect was modest and it did not systematically affect any of the other variables, the decision was made to model the data across Experiments 1 and 2 with a single pooled dataset. The session order factor was retained in the model to account for this variance in performance, resulting in a simplified model that can be expressed as:

$$W \sim Prop + Bg + Env + Prd + Age + Ord + (Prop + Bg + Env + Prd \mid Subj).$$

As described in Experiment 1, additional factors and interactions were included in the model and evaluated for significance using model testing. A combination of step-wise factor addition and elimination was performed to optimize model fit for the design-level factors (*Prop, Bg, Age, Env, Prd,* and *Age)* and their associated interactions. Hypotheses related to the effects of periodicity cues predicted a significant interaction between periodicity and age, such that older adults were predicted to benefit less from periodicity cues than younger adults. Periodicity was also hypothesized to interact with background, such that periodicity cues were predicted to provide additional benefit with a competing talker compared to quiet. An interaction between periodicity and envelope cues was used to evaluate the extent to which these cues provided redundant, additive, or synergistic benefit to speech recognition. Item-level factors (lexical characteristics, sentence length and keyword position, and talker sex) and subject-level factors (education level, hearing sensitivity, and cognitive abilities) were added to the model and tested for significance to determine if these cues influenced the benefit associated with periodicity cues. Due to the large number of factors, item-level factors and subject-level factors were explored separately rather than in a single omnibus model. Item-level and subject-level effects observed in Experiment 1 were retested for significance to determine if the effects remained consistent in the larger dataset that included periodicity cues.

Standard estimates, standard error, and *z*-statistics for each factor, including split factors from *post-hoc* models, can be found in Appendix 2.

III.C. Results

III.C.1. Design-level factors

Keyword recognition (rau) is plotted for the quiet and competing talker backgrounds in Figures 8 and 9, respectively. These data include silence-interrupted and envelope-filled conditions from Experiment 1, as well as the periodicity-filled and envelope-plus-periodicity-filled conditions of Experiment 2.

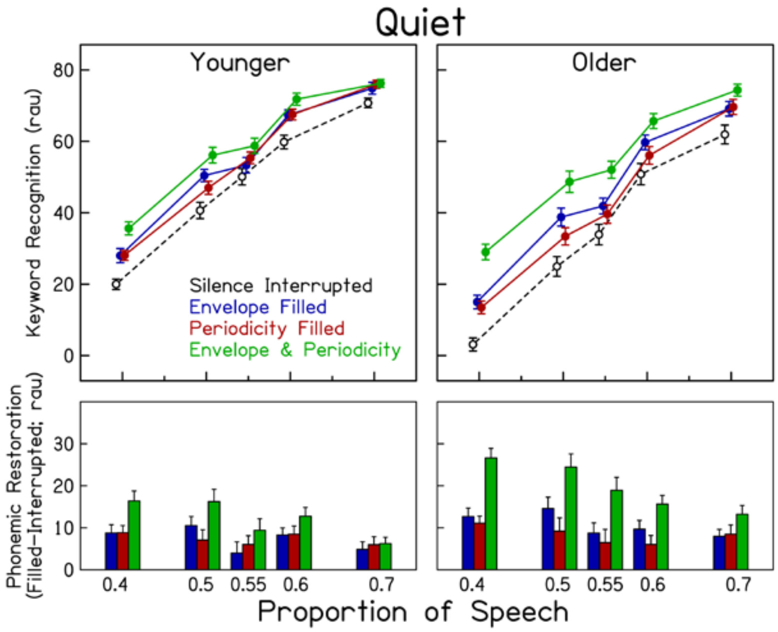

Figure 8: Top panels: Keyword recognition (rau) in quiet background plotted as a function of proportion of speech with interruption type as the parameter. Data from younger adults are plotted on the left panels and from older adults are plotted on the right. Lower panels: Phonemic restoration, or the difference in keyword recognition between silence-interrupted and various filled sentences, is plotted as a function of proportion of speech with interruption type as the parameter.

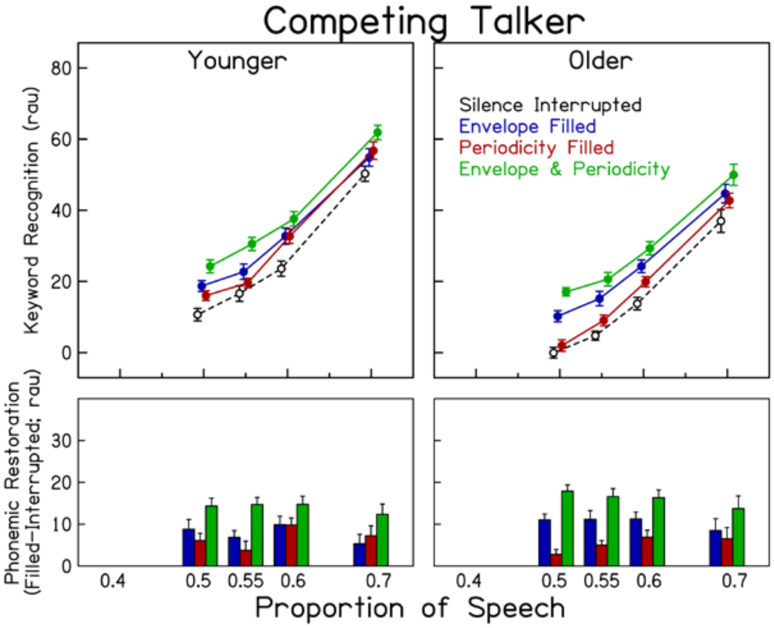

Figure 9: Top panels: Keyword recognition (rau) in competing talker background plotted as a function of proportion of speech with interruption type as the parameter. Data from younger adults are plotted on the left panels and from older adults are plotted on the right. Lower panels: Phonemic restoration, or the difference in keyword recognition between silence-interrupted and various filled sentences, is plotted as a function of proportion of speech with interruption type as the parameter.

A GLMM was constructed to quantify the relative contributions of the design-level factors and their interactions to keyword recognition. Model testing confirmed significant improvements in the fit of the model with the addition of each design-level factor (χ^2_{Prop}=205.35, $p \leq 0.001$; χ^2_{Bg}=173.03, $p \leq 0.001$; χ^2_{Env}=103.65, $p \leq 0.001$; χ^2_{Prd}=55.66, $p \leq 0.001$; χ^2_{Age}=22.08, $p \leq 0.001$). Modeling results were generally similar to Experiment 1: younger adults significantly outperformed older adults (β_{Age}=-0.27, z=-5.42, $p<0.001$); keyword recognition improved significantly with increasing proportion of speech (β_{Prop}=1.00, z=77.60,

$p<0.001$); and keyword recognition declined significantly with a competing talker ($\beta_{Bg}=-0.75$, $z=-53.05$, $p<0.001$). Keyword recognition improved significantly with envelope cues ($\beta_{Env}=0.24$, $z=20.82$, $p<0.001$) and with periodicity cues ($\beta_{Prd}=0.16$, $z=10.81$, $p<0.001$). Several two-way interactions between the design-level factors improved the model fit ($\chi^2_{Prop*Bg}=177.55$, $p\leq0.001$; $\chi^2_{Prop*Age}=18.83$, $p\leq0.001$; $\chi^2_{Bg*Age}=7.04$, $p\leq0.01$; $\chi^2_{Prop*Env}=96.31$, $p\leq0.001$; $\chi^2_{Prop*Prd}=96.31$, $p\leq0.001$; $\chi^2_{Env*Age}=72.37$, $p\leq0.001$; $\chi^2_{Env*Bg}=31.82$, $p\leq0.001$), as well as 3 three-way interactions ($\chi^2_{Age*Env*Prop}=8.86$, $p\leq0.01$; $\chi^2_{Age*Prd*Prop}=6.07$, $p\leq0.05$; $\chi^2_{Bg*Env*Prop}=6.70$, $p\leq0.01$). *Post-hoc* models with split factors were constructed to explore these interaction terms (see next section). Additional two-way, three-way, and four-way interactions were tested for significance using model testing, but none of these additional interaction terms significantly improved the fit of the model ($\chi^2<3.30$, *ns* in all cases).

As in Experiment 1, proportion of speech significantly interacted with background ($\beta_{Prop*Bg}=0.11$, $z=11.45$, $p<0.001$). To explore this interaction, separate factors were generated for speech proportion in quiet and with a competing talker (*Prop_Q* and *Prop_CT*); a model was constructed in which these two factors replaced proportion of speech and its interaction with background (*Prop* and *Prop*Bg*). Results of this split-factor model indicated that proportion was a stronger predictor of keyword recognition with a competing talker ($\beta_{Prop\_CT}=1.10$, $z=65.88$, $p<0.001$) than in quiet ($\beta_{Prop\_Q}=0.88$, $z=82.28$, $p<0.001$). This finding is in agreement with results from Experiment 1;

performance declined more precipitously with decreasing speech proportion with a competing talker than in quiet.

The interaction between proportion of speech and age group observed in Experiment 1 also remained significant in Experiment 2 ($\beta_{Prop*Age}$=0.06, z=4.91, p<0.001). Separate factors were generated for effects of speech proportion for younger and older adults (*Prop*_Y and *Prop_CT*); a model was generated using these factors in place of proportion and the interaction with age (*Prop* and *Prop*Age*). Similar to Experiment 1, results of the split-factor model indicated that proportion was a stronger predictor of keyword recognition among older adults (β_{Prop_O}=1.00, z=75.48, p<0.001) than younger adults (β_{Prop_Y}=0.89, z=71.90, p<0.001).

In contrast to Experiment 1, the interaction between age and background was significant in Experiment 2 (β_{Bg*Age}=-0.04, z=-2.78, p<0.01). The interaction was explored with separate background factors for younger and older adults (*Bg_Y* and *Bg_O*). These factors replaced background and the interaction with age (*Bg* and *Bg*Age*) in a split-factor model. Results indicated that declines in keyword recognition with the competing talker were larger for older adults (β_{Bg_O}=-0.74, z=-67.58, p<0.001) than younger adults (β_{Bg_Y}=-0.68, z=-66.99, p<0.001). This finding is consistent with an age-related decline in speech segregation, but the size of the effect was modest and only became significant once data were pooled across both experiments.

Envelope cues significantly interacted with speech proportion ($\beta_{Prop*Env}$=-0.10, z=-9.80, p<0.001). This interaction was explored with separate speech proportion factors for sentences with and without envelope cues (*Prop_Env* and *Prop_xEnvx*). A split-factor model was constructed in which these factors replaced proportion and its interaction with envelope cues (*Prop* and *Prop*Env*). The results of this split-factor model indicated a more precipitous decline in keyword recognition with decreasing proportion for sentences without envelope cues (β_{Prop_xEnvx}=1.03, z=77.47, p<0.001) than for sentences with envelope cues (β_{Prop_Env}=0.87, z=69.96, p<0.001). This suggests that envelope cues provided listeners with a supportive scaffold that facilitated speech recognition with shorter glimpses.

Similarly, periodicity cues significantly interacted with speech proportion ($\beta_{Prop*Prd}$=-0.04, z=-4.45, p<0.001). This interaction was explored with separate proportion factors for sentences with and without periodicity cues (*Prop_Prd* and *Prop_xPrdx*). A split-factor model was constructed in which these factors replaced proportion and its interaction with periodicity cues (*Prop* and *Prop*Prd*). The results of this split-factor model were similar to the interaction with envelope cues: the decline in keyword recognition with decreasing proportion was greater for sentences without periodicity cues (β_{Prop_xPrdx}=0.98, z=76.89, p<0.001) than for sentences with periodicity cues (β_{Prop_Prd}=0.91, z=74.18, p<0.001). Thus, periodicity cues also provided a source of support that assisted listeners with recognition of speech with shorter glimpses.

Envelope cues significantly interacted with age ($\beta_{Env*Age}$=0.06, z=5.91, p<0.001). This interaction was explored with separate envelope factors for younger and older adults (*Env_Y* and *Env_O*). A split-factor model was constructed in which these factors replaced envelope cues and the interaction with age (*Env* and *Env*Age*). Results of this split-factor model indicated that envelope cues were more beneficial for older adults (β_{Env_O}=0.26, z=25.75, p<0.001) than younger adults (β_{Env_Y}=0.15, z=15.90, p<0.001). Interestingly, the interaction between age and periodicity cues did not reach significance ($\chi^2_{Age*Prd}$=2.42, *ns*). Thus, older adults were able to take greater advantage of envelope cues to facilitate speech recognition than younger adults, but younger and older adults did not differ in their ability to use periodicity cues.

Envelope cues also significantly interacted with background (β_{Env*Bg}=0.05, z=5.63, p<0.001). This interaction was explored with separate envelope factors in quiet and with a competing talker (*Env_Q* and *Env_CT*). A split-factor model was constructed in which these factors replaced envelope cues and the interaction with background (*Env* and *Env*Bg*). Results of this split-factor model indicated that envelope cues were more beneficial with a competing talker (β_{Env_CT}=0.22, z=19.75, p<0.001) than in quiet (β_{Env_Q}=0.19, z=21.65, p<0.001). Once again, a similar interaction between background and periodicity did not reach significance (χ^2_{Bg*Prd}=0.23, *ns*). These results suggest that envelope cues (not periodicity cues) provided a benefit to speech segregation with a competing talker.

The interaction between envelope cues and periodicity cues was tested for significance to determine whether the combined effects of these two cues were redundant, additive, or synergistic. This interaction term did not improve model fit ($\chi^2_{Env*Prd}$=0.21, *ns*), indicating that the combined effects were not redundant or synergistic. Rather, both cues provided separate benefits to speech recognition and these benefits were additive when both cues were available.

A three-way interaction was observed between age, envelope cues, and speech proportion ($\beta_{Prop*Env*Age}$=-0.03, z=-2.98, p<0.01). This interaction was explored by generating separate envelope-by-proportion interaction terms for younger and older adults (*Prop*Env_Y* and *Prop*Env_O*). A split-factor model was constructed in which these two factors replaced the envelope-by-proportion interaction and the three-way interaction with age (*Prop*Env* and *Prop*Env*Age*). Results of this split-factor model indicated that the interaction between envelope cues and proportion was more pronounced in older adults ($\beta_{Prop*Env\_O}$=-0.10, z=-7.61, p<0.001) than in younger adults ($\beta_{Prop*Env\_Y}$=-0.05, z=-3.98, p<0.001). Thus, the supportive scaffold that envelope cues provided for speech recognition with shorter glimpses was more beneficial for older adults than younger adults.

A three-way interaction was also observed between age, periodicity cues, and speech proportion ($\beta_{Prop*Prd*Age}$=-0.02, z=-2.46, p<0.05). This interaction was explored by generating separate periodicity-by-proportion interaction terms for younger and older adults (*Prop*Prd_Y* and *Prop*Prd_O*). A split-factor model was constructed in which these two factors replaced the periodicity-by-proportion interaction and the three-way interaction with age (*Prop*Prd* and *Prop*Prd*Age*). Results of this split-factor model indicated that the interaction between periodicity cues and proportion was nonsignificant for younger adults ($\beta_{Prop*Prd\_Y}$=-0.02, *z*=-1.38, *ns*). Thus, only older adults benefited from periodicity cues as a supportive scaffold for recognition of speech with shorter glimpses ($\beta_{Prop*Prd\_O}$=-0.06, *z*=-4.68, *p*<0.001). This result was unexpected and suggests that older adults may benefit from any supportive cues to facilitate recognition of short glimpses of speech.

The final significant three-way interaction was between background, envelope cues, and proportion ($\beta_{Prop*EnvBg}$=-0.03, *z*=-2.59, *p*<0.01). This interaction was explored by generating separate envelope-by-proportion interactions in quiet and with a competing talker (*Prop*Env_Q* and *Prop*Env_CT*). A split-factor model was constructed in which these two factors replaced the envelope-by-proportion interaction and the three-way interaction with background (*Prop*Env* and *Prop*Env*Bg*). Results of this split-factor model indicated that the interaction between envelope cues and proportion was more pronounced with a competing talker ($\beta_{Prop*Env\_CT}$=-0.11, *z*=-6.70, *p*<0.001) than in quiet ($\beta_{Prop*Env\_Q}$=-0.06, *z*=-5.88, *p*<0.001). This suggests that the structured pattern of modulation provided by envelope cues was particularly beneficial for connecting short glimpses of speech in the competing talker background.

III.C.2. Item-level factors

III.C.3.a. Lexical characteristics

Lexical characteristics described in Experiment 1 (word frequency, neighborhood density, and biphone probability) were retested for significance in the GLMM. Model testing confirmed significant contributions of each lexical characteristic (χ^2_{WF}=1481.5, *p*≤0.001; χ^2_{ND}=422.58, *p*≤0.001; χ^2_{BP}=37.47, *p*≤0.01), as well as the interaction between word frequency and envelope cues (χ^2_{Env*WF}=7.31, *p*≤0.01). Modeling results were similar to those reported in Experiment 1. Keyword recognition was better for more commonly used words than for less common words (β_{WF}=0.27; *z*=38.38; *p*≤0.001) and recognition was better for keywords with more common phoneme sequences (β_{BP}=0.04; *z*=6.13; *p*≤0.001). Recognition was poorer for keywords from more dense neighborhoods (β_{ND}=-0.15; *z*=-20.32; *p*≤0.001). The significant interaction between word frequency and envelope cues (β_{Env*WF}=-0.02; *z*=-2.70; *p*≤0.01) was explored with a *post-hoc* model with separate envelope factors for more commonly used keywords (i.e., normalized *WF*>0; *Env_hiWF*) and less common keywords (i.e., normalized *WF*<0; *Env_loWF*). Results of the split-factor model indicated greater improvements with envelope cues for less common keywords (β_{Env_loWF}=0.25; *z*=20.37; *p*≤0.001) than more common keywords (β_{Env_hiWF}=0.21; *z*=13.96; *p*≤0.001). Additional interactions, including interactions with periodicity, were tested for significance but did not improve the fit of the model (χ^2<2.09, *ns* in all cases). These results suggest that whereas the lexical characteristics of the keywords influenced recognition, these effects were generally independent of the effects of age and temporal cues, with the exception of envelope cues providing a greater benefit for recognition of less common words.

III.C.3.b. Sentence length and keyword position

Sentence length (*nWords*) and keyword position (*Pos*) factors described in Experiment 1 were retested for significance in the GLMM. Model testing confirmed significant contributions of both factors (χ^2_{nWords}=434.60, $p \leq 0.001$; χ^2_{Pos}=311.28, $p \leq 0.001$). As in Experiment 1, keyword recognition was poorer for sentences with more keywords than for shorter sentences (β_{nWords}=-0.15; z=-20.24; $p \leq 0.001$), and recognition improved over the course of the sentence (β_{Pos}=0.13; z=17.60; $p \leq 0.001$). The three interactions that were significant in Experiment 1 remained significant in Experiment 2 ($\chi^2_{Pos*Age}$=40.62, $p \leq 0.001$; $\chi^2_{nWords*Bg}$=155.29, $p \leq 0.001$; χ^2_{Pos*Bg}=102.97, $p \leq 0.001$) and were not re-evaluated with *post-hoc* models. Two additional interactions with keyword position also significantly improved model fit ($\chi^2_{Pos*Env}$=4.58, $p \leq 0.05$; $\chi^2_{Pos*Prd}$=7.09, $p \leq 0.01$) and these new interactions were explored with *post-hoc* models.

The interaction between keyword position and envelope cues ($\beta_{Pos*Env}$=0.01, z=2.14, p<0.05) was explored with a *post-hoc* model with separate keyword position factors for sentences with and without envelope cues (*Pos_Env* and *Pos_xEnvx*); results revealed that the improvement in keyword recognition over the course of the sentence was greater for sentences with envelope cues (β_{Pos_Env}=0.13; z=13.15; $p \leq 0.001$) than without envelope cues (β_{Pos_xEnvx}=0.09; z=9.14; $p \leq 0.001$). This suggests that continuous envelope cues facilitated a consistent gradual improvement in recognition of keywords over the course of the sentence.

The interaction between keyword position and periodicity cues ($\beta_{Pos*Prd}$=-0.02, z=-2.71, p<0.01) was explored with separate keyword position factors for sentences with and without periodicity cues (*Pos_Prd* and *Pos_xPrdx*); results of this split-factor model indicated that the improvement in recognition over the course of the sentence was greater without periodicity cues (β_{Pos_xPrdx}=0.13; z=12.84; $p \leq 0.001$) than with periodicity cues (β_{Pos_Prd}=0.09; z=9.61; $p \leq 0.001$). Thus, periodicity cues may have facilitated more sporadic recognition of intelligible keywords, rather than the gradual time-varying improvements noted with envelope cues.

III.C.3.c. Talker sex

Effects and interactions related to talker sex were all consistent with results of Experiment 1. As a single fixed effect, talker sex did not improve model fit (χ^2_{Sex}=0.63, *ns*), but interactions were significant between talker sex and age ($\chi^2_{Sex*Age}$=128.33, $p \leq 0.001$), talker sex and background (χ^2_{Sex*Bg}=35.88, $p \leq 0.001$), and the three-way interaction between age, background, and talker sex ($\chi^2_{Sex*Bg*Age}$=6.11, $p \leq 0.05$). Modeling results were consistent with Experiment 1. The interaction between talker sex and age ($\beta_{Sex*Age}$=-0.08; z=-11.23; $p \leq 0.001$) was driven by better recognition of female talkers among younger adults (β_{Sex_Y}=0.08; z=8.66; $p \leq 0.001$) and better recognition of male talkers among older adults (β_{Sex_O}=-0.07; z=-6.99; $p \leq 0.001$). The interaction between talker sex and background (β_{Sex*Bg}=0.04; z=5.99; $p \leq 0.001$) was driven by better recognition of male talkers in quiet (β_{Sex_Q}=-0.04; z=-4.34; $p \leq 0.001$) and better recognition of female talkers with a competing talker (β_{Sex_CT}=0.05; z=4.51; $p \leq 0.001$). Results of the three-way interaction were also generally consistent with Experiment 1 ($\beta_{Sex*Bg*Age}$=-0.02; z=-2.47; $p \leq 0.05$). In the quiet background, younger adults had a modest advantage for recognition of female talkers ($\beta_{Sex\_Q\_Y}$=0.03; z=2.06; $p \leq 0.05$), whereas older adults had greater difficulty with female talkers compared to male talkers ($\beta_{Sex\_Q\_O}$=-0.10; z=-8.01; $p \leq 0.001$). With a competing talker, younger adults demonstrated a considerable advantage for recognition of female talkers compared to males ($\beta_{Sex\_CT\_Y}$=0.15; z=9.86; $p \leq 0.001$), whereas older adults had a slight disadvantage for female talkers ($\beta_{Sex\_CT\_O}$=-0.05; z=-3.08; $p \leq 0.01$).

III.C.3. Subject-level factors

As in Experiment 1, effects of education level, hearing sensitivity, and cognitive abilities were tested for significance as subject-level factors in the GLMM. Education level significantly improved model fit (χ^2_{Edu}=5.67, $p \leq 0.05$) and modeling results indicated better keyword recognition with increasing years of education (β_{Edu}=0.10; z=2.47; $p \leq 0.05$). Hearing sensitivity was evaluated with the 4 age-residualized PTA measures described in Experiment 1 (NPTA, BPTA, LFPTA, and HFPTA). None of these PTA measures significantly improved the fit of the model (χ^2<2.71, *ns*). Age-residualized cognitive measures described in Experiment 1 were also tested for significance (Connections, Peg Board, Reading Span, Stroop, and TRT), and model testing revealed that only the Connections score significantly improved model fit ($\chi^2_{Connections}$=11.12, $p \leq 0.001$). Modeling results indicated better keyword recognition among subjects with faster processing speed ($\beta_{Connections}$=0.15; z=3.59; $p \leq 0.001$). Other cognitive measures and interactions between cognitive measures and design-level factors were not predictive of keyword recognition (χ^2<4.90, *ns*), including TRT, which was significant in the smaller dataset from Experiment 1.

III.D. Discussion

This study investigated the effects of age on the use of periodicity cues for recognition of interrupted speech in quiet and with a competing talker. Data were pooled across Experiments 1 and 2 in order to determine the relative contributions of envelope and periodicity cues to speech recognition as well as their combined benefit when both cues were available. As in Experiment 1, older adults performed poorer than younger adults across all speech conditions and variance in hearing sensitivity among younger and older adults did not significantly predict keyword recognition. Performance declined as the proportion of speech decreased and performance was poorer with a competing talker than in quiet. The additional data on periodicity cues provided further explanation for age-related changes in speech recognition and, with very few exceptions, results from Experiment 1 were consistent with results from the larger dataset in Experiment 2.

One notable exception was the effect of session order, which was not significant in Experiment 1 but was significant in Experiment 2. The most likely reason for this change is that data from Experiment 1 were collected in a single session, which was randomly assigned to be the first or second session for each subject. Thus, session order was essentially a between-subject variable in Experiment 1; results indicated that performance of subjects who completed testing in session 1 did not differ from those who completed testing in session 2. The analysis for Experiment 2 included data collected from both sessions for each subject and thus session order was a within-subject variable in the larger analysis. These results indicated a modest, but significant, improvement in keyword recognition from the first to the second session. As the magnitude of this effect was small, and was consistent across conditions and age groups, the data were pooled across the two sessions and the session order factor was retained in the model to account for the increased variance in the data.

III.D.1. Relative contributions of envelope and periodicity cues

Envelope and periodicity cues provided separate and additive contributions to recognition of interrupted speech. Consistent with recent work by Oh and colleagues (2016), speech recognition was best when both cues were available. Performance improved when either cue was available and the additional cues facilitated recognition with shorter glimpses of speech. More specifically, modeling results indicating a greater benefit associated with envelope cues than periodicity cues both in quiet and with a competing talker. The envelope is naturally continuous in an uninterrupted sentence, and so the additional envelope cues provided here likely restored the natural continuity of the envelope of the interrupted sentences. In contrast, periodicity is naturally intermittent in an uninterrupted sentence, occurring only during voiced speech segments and vowels. As such, the additional periodicity cues provided here may have created an unnatural continuity of periodicity information through segments of speech that would otherwise be aperiodic, such as unvoiced consonants. These spurious periodic segments can result in poorer speech recognition compared to an analogous aperiodic stimulus (Steinmetzger & Rosen, 2015). Aperiodic noise however, contains other spurious cues. Random fluctuations in the amplitude of noise can disrupt recognition of envelope modulations in glimpses of speech (Stone et al., 2012). Similar modulation masking effects likely occurred when silent intervals were filled with envelope-modulated noise, though these effects were outweighed by the improvement in recognition associated with phonemic restoration. Envelope-modulated pulse trains carried considerably fewer random fluctuations than noise and may have resulted in less modulation masking. As such, the improved recognition of envelope-plus-periodicity filled sentences may also reflect a release from modulation masking, relative to the envelope-filled condition.

With a competing talker, periodicity cues were hypothesized to provide an additional benefit to speech recognition. Contrary to this hypothesis, the benefit associated with continuous periodicity cues was equivalent in quiet and with a competing talker and envelope cues provided additional benefit for segregation of speech, similar to results observed in Experiment 1. These results were further supported by the observation that envelope cues provided the greatest benefit in the most difficult listening conditions, that is, recognition of short glimpses of speech with a competing talker. Thus, it appears that continuity of the temporal envelope allowed listeners to more effectively track glimpses of target speech in the presence of a competing talker. Periodicity cues provide no additional benefit with a competing talker, suggesting that periodicity cues were either redundant with existing periodicity in the speech glimpses, or were otherwise unnecessary for speech segregation.

Keyword recognition generally improved over the course of the sentence, but the extent of this improvement was affected by envelope and periodicity cues within the interrupted portions. Sentences with envelope cues demonstrated steady improvements in keyword recognition over the course of the sentence, whereas the same effect was not observed with periodicity cues. Rather, the presence of periodicity cues decreased the association between recognition and keyword position. Envelope cues may have provided a predictable pattern that helped listeners track the target talker throughout the sentence, whereas periodicity cues provided moment-to-moment improvements in recognition of individual keywords, especially when longer vowel portions of a keyword were removed by the sentence interruption. These vowel segments may have benefited more from a periodicity-based filler signal, which more closely resembles a vowel than envelope-modulated noise or silence. Nevertheless, these cases would have occurred at random points in sentences and thus would not have contributed to the model’s ability to predict recognition based on keyword position.

III.D.2. Effects of age

Similar to Experiment 1, older adults demonstrated greater difficulty than younger adults connecting short glimpses of speech across time. Performance declined precipitously for older adults with decreasing proportion of speech for all sentence types. The addition of envelope and/or periodicity cues helped facilitate glimpsing in older adults, but these benefits did not fully compensate for the effects of age. These results largely confirm the findings from Experiment 1, which indicated that age-related declines in glimpsing account for the largest portion of variance in recognition of interrupted speech.

In contrast to Experiment 1, older adults demonstrated a larger decline in speech recognition than younger adults when speech was presented with a competing talker. This finding is consistent with previous studies showing poorer recognition of speech by older adults compared to younger adults in competing talker backgrounds (e.g., Helfer & Freyman, 2008; Lee & Humes, 2012). However, the effect was modest and was likely revealed due to greater sensitivity in the larger model. The additional data from Experiment 2 also included higher keyword recognition scores, which may have increased overall performance such that floor effects no longer obscured the effect of age on speech segregation. Other studies have shown a prolonged time course of speech segregation in older adults (Ben-David et al., 2012; Ezzatian et al., 2015). In Experiment 2, recognition improved over the course of the sentence with a competing talker but these effects were equivalent in younger and older adults. Taken together, these results suggest that effects of age on speech segregation are relatively small in comparison to the more pronounced age-related declines observed for glimpsing.

Older adults were expected to benefit less from periodicity cues than younger adults. This hypothesis was supported by studies showing age-related declines in periodicity coding (Clinard & Tremblay, 2013; Snyder & Alain, 2005) and declines in sentence recognition when sentences were temporally “jittered” to simulate poor periodicity coding (Pichora-Fuller et al., 2007). In Experiment 2, younger and older adults achieved equivalent benefit from periodicity cues.

Furthermore, the extent to which periodicity cues facilitated glimpsing for the shortest speech segments was *greater* for older adults than younger adults. Thus, any age-related declines in periodicity coding had a negligible effect on older adults' ability to use a continuous F0 contour to facilitate speech recognition and glimpsing.

III.D.3. Talker sex and additional factors

The effect of talker sex was similar to that observed in Experiment 1; larger age-related declines in speech recognition were observed for female talkers than male talkers. An intuitive explanation for this effect is that age-related declines in periodicity coding may have a greater effect on recognition of female talkers than male talkers, due to the faster rate of periodicity cues in female voices. However, effects of periodicity cues were relatively modest in this study. If poor coding or use of periodicity information in female voices was responsible for the decline in performance in older adults, then the additional periodicity cues provided by the pulse trains should have provided a greater benefit on sentences with female talkers. This was not supported by the data, which revealed that periodicity cues provided an equivalent benefit for recognition of male and female talkers. As such, the source of the age-related decline in recognition of female talkers remains unclear and serves as an important future direction for research on the effects of age on speech recognition in realistic listening environments.

In Experiment 1, the TRT was highly predictive of recognition of silence-interrupted speech. In the larger dataset containing periodicity-filled and envelope-plus-periodicity-filled sentences, the TRT did not significantly predict keyword recognition. This suggests that the TRT is uniquely sensitive to recognition of silence-interrupted speech. The benefits listeners achieved from additional envelope and/or periodicity cues improved keyword recognition and reduced the extent to which performance could be predicted based on an amodal ability to make use of partial linguistic information, as measured by the TRT.

In contrast, speed of processing remained a significant predictor of keyword recognition in the larger dataset. This suggests that speed of processing may be a more general ability required for incorporating segments of speech across gaps of missing information. Whereas envelope and periodicity cues may provide a beneficial scaffold that improves sentence recognition, the association between processing speed and overall performance persists. Thus, processing speed may be an important cognitive ability in realistic listening environments where speech is masked by a fluctuating background, leaving only short audible glimpses that must be incorporated efficiently by the listener for adequate recognition.

III.E. Conclusions

Experiment 2 assessed the relative contributions of envelope and periodicity cues to recognition of interrupted speech by younger and older adults in quiet and with a competing talker. Envelope cues provided a greater benefit than periodicity cues for keyword recognition, as well as for glimpsing and speech segregation. Benefits of envelope and periodicity cues were additive when both cues were available. Though older adults were expected to benefit less than younger adults from periodicity cues, results indicated that older adults benefit at least as much from these faster rate cues as younger adults. Although effects of age on speech segregation were observed, they were relatively small. Taken together, these results suggest that age-related declines in recognition of interrupted speech are best explained by difficulty with glimpsing and that older adults can partially compensate for these declines when supportive envelope and periodicity cues are available. Subject-level differences in speed of processing contributed to individual differences in performance, suggesting that speech recognition may depend on the extent to which a listener can quickly and efficiently incorporate glimpses of speech into a coherent representation of the message.

IV. Experiment 3: Object Selection Based on Attention and Expectations of Voice Characteristics

IV.A. Introduction

Attention plays an intuitive role in speech perception in multitalker environments. When several sources of speech are present, attention must be directed to a particular talker, and competing talkers must be ignored to avoid processing irrelevant speech information (Zion Golumbic et al., 2013). The process of selecting a particular talker as the focus of attention (object selection) can be distinguished from lower level processes that facilitate segregation of speech glimpses from competing talkers and connection of those glimpses across time into a single coherent percept (object formation). While object formation and object selection can be viewed as distinct processes, few studies have attempted to isolate effects related to object selection from those of object formation (Ihlefeld & Shinn-Cunningham, 2008). Experiments 1 and 2 evaluated the effects of age and the role of envelope and periodicity cues in specific components of object formation (glimpsing, speech segregation, phonemic restoration). Experiment 3 evaluated the effects of age on object selection and its contribution to speech recognition in realistic listening environments.

Several studies have demonstrated that older adults experience greater difficulty than younger adults understanding speech in backgrounds with competing talkers (Ben-David et al., 2012; Duquesnoy, 1983, Festen & Plomp, 1990; Helfer & Freyman, 2008; Humes et al., 2006; Lee and Humes, 2012; Li et al., 2004; Rajan & Cainer, 2008; Tun et al., 2002). The results of these studies are consistent with several different interpretations for the observed effects of age and thus it remains unclear to what extent these age-related declines reflect difficulty with object formation or object selection. Some results demonstrate an age-related increase in masker errors and distraction by competing speech, which are suggestive of difficulty with object selection (Helfer & Freyman, 2008; Humes et al., 2006; Tun et al., 2002). Other results demonstrate a prolonged time course for speech segregation among older adults, which is suggestive of declines in object formation (Ben-David et al., 2012; Ezzatian et al., 2015), similar to the interpretation of age-related declines in glimpsing observed in Experiments 1 and 2. However, declines in object formation and object selection may not be mutually exclusive and therefore the potential effects of age on object selection were explored separately in Experiment 3.

Several experimental methods have been developed to study the role of attention on speech recognition. Providing listeners with *a priori* knowledge of a target's voice characteristics or spatial location can greatly improve speech recognition, as compared to conditions in which similar information on the target is provided after stimulus presentation (Humes et al., 2006; Ihlefeld & Shinn-Cunningham, 2008; Kidd et al., 2005). Similar declines in speech recognition are observed with other dual-task paradigms in which listeners are required to perform a concurrent secondary task (Gallun, Mason, & Kidd, 2007; Helfer, Chevalier, & Freyman, 2010). Older adults typically demonstrate greater difficulty with dual-tasks and divided attention than younger adults and these effects have been associated with age-related declines in working memory capacity (Humes et al., 2006; Verhaegen, Steitz, Sliwinski, & Cerella, 2003) as well as attentional control (Bier, Lecavalier, Malenfant, Peretz, & Belleville, 2017). However, the extent to which age-related declines in speech recognition can be explained by declines in object selection and the ability to focus attention on an expected voice remains unclear.

Another method of assessing effects of attention and stimulus expectations is the “probe-signal” method (Dai et al., 1991; Greenberg & Larkin, 1968; Sharf et al., 1987; Schlauch & Hafter, 1991). In their seminal paper, Greenberg and Larkin trained listeners to detect a 1.0 kHz tone in broadband noise. For a small subset of trials, the tone was replaced with a “probe” tone at a different frequency. Detection of probe tones declined as a function of the log-frequency distance from 1.0 kHz, despite all tones being presented at levels adjusted to be equally detectable. The results revealed a characteristic band-pass filter shape (the “attentional filter”), which described the benefit of focused attention for detection of tones in the expected frequency range around 1.0 kHz. Subsequent applications of the probe-signal method demonstrated that attentional filtering occurs for several other acoustic dimensions, including duration and temporal structure (Dai & Wright, 1995; White & Carlyon, 1997). Whereas this paradigm has been successful in describing the effects of focused attention on non-speech auditory detection tasks, it has not yet been adapted for assessing the contributions of attention and stimulus expectations to speech recognition.

In order to adapt the probe-signal method for a speech recognition task, talker sex was selected as a suitable dimension for influencing listeners' expectations of the target. Perception of talker sex can be manipulated in a systematic way using signal processing methods to alter a talker's F0 and spectral envelope (Darwin et al., 2003; Mackersie et al., 2011); these cues have been used to facilitate object selection in numerous studies of speech recognition with competing talkers (Duqesnoy, 1983; Festen & Plomp, 1990; Helfer & Freyman, 2008; Humes et al., 2006; Mackersie et al., 2011). These two acoustic cues are also known to facilitate strong speech segregation effects (Brokx and Nooteboom, 1982; Gaudrain, Grimault, Healy, & Béra, 2008), suggesting important contributions of F0 and spectral envelope to object formation. As such, the relative contributions of these voice cues to object selection and object formation are unresolved in many speech recognition studies. The probe-signal method is well suited to tease apart these effects and determine the extent to which speech recognition difficulties of older adults reflect a decline in object selection.

Experiment 3 used a modified version of the probe-signal method to test the hypothesis that older adults are less able than younger adults to focus attention on an expected voice to facilitate speech recognition in a realistic listening environment. Male and female sentence pairs were processed for a standardized difference in F0 and spectral envelope to control for the effects of these cues on speech segregation. An equal number of trials contained male vs. female target talkers, and listeners were trained to identify the target based on a “cue phrase,” spoken by the target talker in quiet prior to each two-talker mixture. On the majority of trials (“standard trials”), the F0 and spectral envelope of the cue phrase were identical to the target, and listeners were expected to benefit from focused attention to these voice features to facilitate object selection and speech recognition. For a small subset randomly occurring “probe” trials, the F0 and spectral envelope of the cue phrase were parametrically shifted towards the competing talker’s voice. Performance on probe trials was expected to decline as the cue phrase was shifted further from the target talker’s voice characteristics. The extent of this decline represents the benefit associated with focused attention to the expected voice features of the target on standard trials. As such, age-related declines in the use of focused attention for object selection would be indicated by large differences in performance between younger and older adults on standard trials and more similar performance between groups on probe trials.

Another method of assessing failures of object selection is to measure “masker errors” or trials where the listener incorrectly reports words from the competing talker sentence. Analyses of masker errors have been used in several other studies of speech recognition with competing talkers and are particularly useful for attributing declines in performance to listeners selecting the wrong talker as the focus of attention (Helfer & Freyman, 2008; Ihlefeld & Shinn-Cunningham, 2008; Lee & Humes, 2012). Results often reveal an increased number of masker errors in responses by older adults, relative to younger adults, providing further support for the hypothesis that declines in object selection contribute to age-related speech recognition difficulties. In Experiment 3, several words in competing talker sentences were selected prior to testing to serve as “masker keywords,” and participant responses were scored for both target keywords correct and masker errors. Masker errors were analyzed to determine if the decline in performance on probe trials could be explained by a greater number of masker errors, and if older adults demonstrated a greater number of masker errors than younger adults.

IV.B. Methods

IV.B.1. Subjects

The same 40 subjects from Experiments 1 and 2 participated in Experiment 3. This experiment used uninterrupted PRESTO sentences that were considerably more intelligible than the interrupted versions presented in Experiments 1 and 2. In order to minimize the effects of sentence familiarity, Experiments 1 and 2 were completed before Experiment 3 for all subjects. Thus, all subjects had the same amount of experience with interrupted versions of the PRESTO sentences at the time of data collection using uninterrupted sentences for Experiment 3.

IV.B.2. Stimulus design and processing

Uninterrupted sentences from the PRESTO lists served as targets and were paired with opposite-sex competing talker sentences from the TIMIT corpus. Target and competing sentences were processed in Praat with the PSOLA algorithm (Moulines & Charpentier, 1990), such that each sentence pair had a standardized difference in F0 and spectral envelope. For each sentence pair, the geometric mean between the male and female F0 was calculated to serve as a midpoint. Next, pitch contours for each sentence were extracted and shift factors were calculated to adjust the male and female F0 to be exactly four semitones above and below the midpoint using the following formulae:

$$\text{Female Shift Factor} = \frac{2^{(+4/12)} * \text{Midpoint}}{\text{Original Female F0}}$$

$$\text{Male Shift Factor} = \frac{2^{(-4/12)} * \text{Midpoint}}{\text{Original Male F0}}$$

The difference in the sign of the exponent ensured that the pitch contour of the female sentence was 4 semitones above the midpoint and the pitch contour of the male sentence was four semitones below the midpoint. By multiplying the pitch contours by these shift factors, the average F0 of the resulting pitch contours differed by exactly eight semitones, while maintaining natural variations in F0 across different male and female talkers. Pilot testing indicated that an 8-semitone difference resulted in performance on standard trials that was free of ceiling effects and an observable decline in performance for probe trials. After processing the mean F0 was 125.2 Hz (*SD*=11.1 Hz) for male talkers and 198.3 Hz (*SD*=17.7 Hz) for female talkers.

Spectral envelopes of the target and competing talkers were manipulated in Praat based on methods described by Darwin et al. (2003). Linear extrapolation of average male/female ratios for the formant and F0 data reported by Peterson and Barney (1952) were used to obtain spectral envelope shift factors (vt) corresponding to a given shift in semitones of F0. For a given sentence, the semitone shift in F0 applied to the pitch contour was used to find the corresponding vt value from Darwin et al. (2003). Then, F0 was multiplied by vt and the duration of the sentence was multiplied by 1/vt. Next, the resulting signal was resampled at the original sampling frequency multiplied by vt. Finally, the signal was saved as a .wav file at the original sampling frequency. The result of this procedure was that the F0 and duration of the sentences were unchanged, but the spectral envelope was scaled by vt. This method of changing the spectral envelope is similar, but not identical, to a true change in vocal tract length and has been used in previous studies of talker sex differences in speech recognition with competing talkers (Darwin et al., 2003; Mackersie et al., 2011). Example spectrograms are displayed in Figure 10 for a single sentence that has been processed for a downward shift in voice features (“male talker”) or an upward shift in voice features (“female talker”).

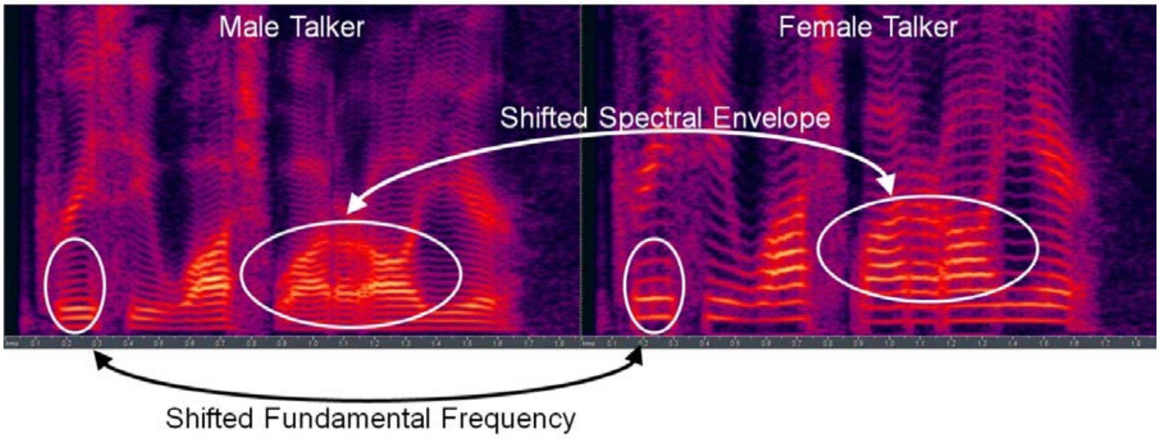

Figure 10: Example spectrograms illustrating a sentence spoken by a male talker (left) and the same sentence processed to sound like a female talker (right). Fundamental frequency is shifted for a higher voice pitch and spectral envelope is broadened such that formants occur in a higher frequency range.

For each target talker, a cue phrase (“greasy wash water all year”) was excised from a standard sentence in the TIMIT corpus spoken by that talker. These cue phrases were processed using similar methods described above to create standard trials and probe trials. Schematic diagrams of the two trial types are displayed in Figure 11. For standard trials (top panel) the cue phrase was processed such that the F0 and spectral envelope matched those of the target talker (i.e., ±4 semitones from the midpoint). For probe trials (bottom panel) the cue phrase was processed such that the F0 and spectral envelope were 1.0, 0.5, or 0 semitones from the midpoint between the target and competing talker. For both trial types, the cue phrase was followed by 1.5 seconds of silence and then the target/competing talker sentence mixture. Thus, the only difference between standard trials and probe trials was the voice characteristics of the cue phrase, which either matched the target (standard trials) or did not match the target (probe trials).

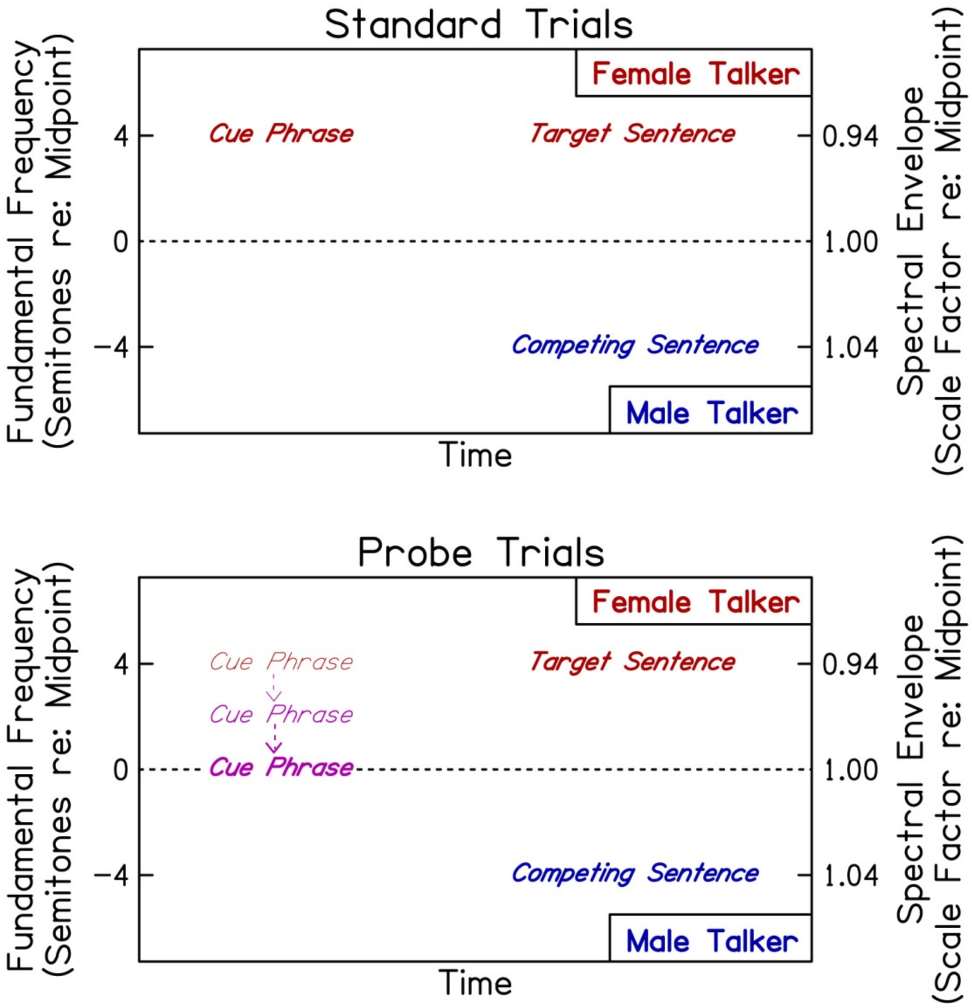

Figure 11: Schematic diagram of standard trials (top panel) and probe trials (bottom panel). All trials began with a cue phrase, followed by 1.5 seconds of silence, and then the target and competing talker mixture. On standard trials, F0 and spectral envelope of the cue phrase matched the target talker exactly. On probe trials, F0 and spectral envelope of the cue phrase were shifted towards the midpoint between the target and competing talker.

Sixteen PRESTO lists were processed as standard trials and 3 PRESTO lists were processed as probe trials, corresponding to the 3 shift conditions (1.0, 0.5, and 0 semitones re: midpoint). The 54 probe trials (3 lists x 18 sentences per list) accounted for 18.75% of all trials and were inserted into the 16 lists of standard trials, such that each list contained 18 standard trials and 3-4 probe trials. Probe trials were randomly assigned to lists with the restriction that each list contained at least 1 probe trial for each shift condition. The order of sentences within each list was randomized for each participant to ensure that probe trials occurred randomly throughout testing and were not presented in a consistent pattern across participants. The last remaining PRESTO list was reserved to be used as a practice list containing only standard trials. The apparatus and calibration procedures were the same as in Experiments 1 and 2, except that target and competing talker mixtures were presented at 78 dB SPL with a 0 dB SNR (75 dB SPL/talker).

IV.B.3. Procedures

Subjects were instructed to use the voice characteristics of the cue phrase to identify the target talker in the subsequent mixture and were not informed about the existence of probe trials. All subjects had previous experience with the competing talker format and cue phrase from Experiments 1 and 2. This facilitated adaptation to the task and use of the sex of the talker in the cue phrase as a reliable target identifier. Testing began with a practice list that contained only standard trials and were not scored. Following the practice list a total of 16 PRESTO lists were presented in random order. Each list consisted of 18 standard trials and 3-4 probe trials presented in random order. Testing was separated into two blocks with a break in between. Speech testing was completed in a single session lasting about two hours.

Important content words in competing talker sentences were selected as masker keywords prior to testing to allow for separate scoring of target keywords correct and masker errors. The number of masker keywords in each sentence did not always match the number of target keywords in the corresponding target sentence, but each list contained an equal number of target and masker keywords (76 target/masker keywords per list). Subject responses on each trial were scored live by the experimenter for target keywords correct and masker errors using a strict scoring rule for both target and masker keywords. Subject responses were also recorded for later confirmation of responses and scoring as needed.

IV.B.4 Statistical approach

Data from standard and probe trials were analyzed with a single item-level logistic regression using a GLMM. Three factors were generated to designate the three probe conditions, corresponding to shifts of 1.0, 0.5, and 0 semitones from the midpoint between the target and competing talker (*Probe1*, *Probe0.5*, and *Probe0*). Standard trials were coded with a negative value for all three probe factors. Probe trials were coded with a positive value for the probe factor corresponding to that shift and a negative value for the other two probe factors. As the number of masker keywords did not match the number of target keywords on each trial, a sentence-level factor was generated to assess the effect of masker errors on keyword recognition. For sentences containing masker errors, the number of masker keywords in the subject's response was divided by the total number of masker keywords for that sentence, resulting in a value between 0 and 1 (*MErr*). Sentences with no masker errors were assigned a *MErr* value of -1. Each target keyword in the sentence was assigned the same *MErr* value describing the proportion of masker errors on that particular sentence. Probe factors and masker errors were used to predict keyword recognition (*W*), the binary dependent variable. Several other predictor variables were also added to the model, including Age group (*Age*), talker sex (*Sex*), sentence length (*nWords*), and keyword position (*Pos*). Preliminary modeling suggested a high degree of subject-level variance in the effects of talker sex and keyword position and so these factors were included in a random subject effects term (*1*|*Subj*). A simplified version of the model can be expressed as:

$$W \sim Age + Sex + nWords + Pos + Probe1 + Probe0.5 + Probe0 + MErr + (Sex+Pos \mid Subj)$$

Each of these factors, as well as interactions and other subject-level factors, were evaluated using model testing to determine if they significantly improved model fit. Interactions were explored using *post-hoc* models with split factors. Standard estimates, standard error, and *z*-statistics for each factor, including split factors from *post-hoc* models, can be found in Appendix 3.

IV.C. Results

Keyword recognition (rau) is plotted in Figure 12 as a function of the F0 shift of the cue phrase (semitones re: midpoint) for younger adults (black) and older adults (gray). Standard trials are plotted as single points on the left and probe trials are plotted as connected symbols on the right. Target keywords correct are plotted with solid lines, and symbols and masker errors are plotted with dashed lines and open symbols. Masker errors on standard trials were rare for younger adults (-15.7 rau) and older adults (-14.6 rau) and are not shown.

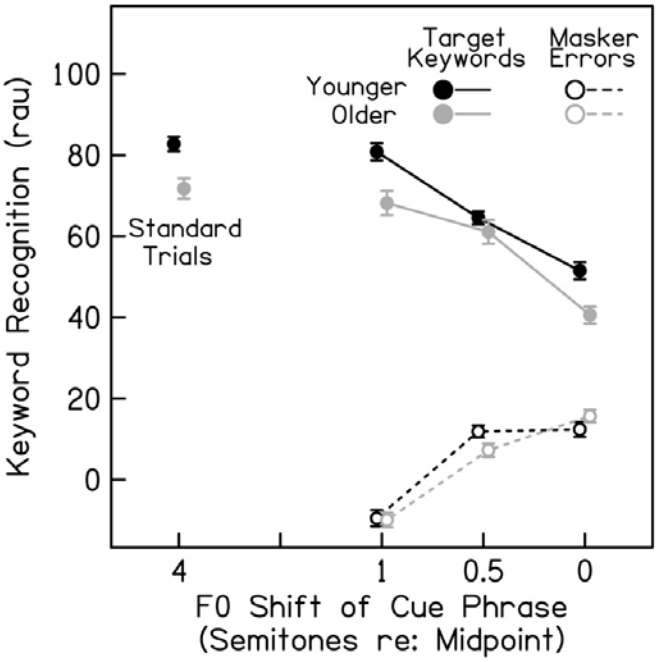

Figure 12: Keyword recognition plotted in rau (solid lines) and masker errors (dotted lines) for younger (black) and older (gray) adults plotted as a function of semitone shift of the cue phrase. Performance is best for standard trials, in which the cue phrase matches the target talker (shift of 0 semitones), and declines for probe trails with increasing shift of the cue phrase. Masker errors were measured in the same rau scale based on selected masker keywords from competing talker sentences. Masker errors increase for probe trials with increasing shift of the cue phrase.

A GLMM was constructed to quantify the effects of the shift in the position of the cue phrase on probe trials and several other predictor variables and interactions. Model testing confirmed that the age factor significantly improved the fit of the model ($\chi^2_{Age}=15.18$, $p \leq 0.001$); modeling results indicated that younger adults significantly outperformed older adults ($\beta_{Age}=-0.24$; $z=-4.30$; $p \leq 0.001$). Model testing also revealed significant improvements in the fit of the model with the addition of probe factors corresponding to the 0.5-semitone shift and the 0-semitone shift ($\chi^2_{Probe0.5}=5.10$, $p \leq 0.05$; $\chi^2_{Probe0}=270.78$, $p \leq 0.001$). Results indicated that keyword recognition declined significantly for probe trials in which the cue phrase was shifted to 0.5 semitones ($\beta_{Probe0.5}=-0.06$; $z=-2.28$; $p \leq 0.05$) and 0 semitones ($\beta_{Probe0}=-0.40$; $z=-17.00$; $p \leq 0.001$) from the midpoint between talkers. The probe factor corresponding to the 1-semitone shift was added to the model and tested for significance, but model testing revealed that this factor did not improve model fit ($\chi^2_{Probe1}=0.56$, *ns*). Thus, shifting the cue phrase to 1 semitone from the midpoint did not have a significant effect on keyword recognition. The proportion of masker errors on each sentence significantly improved the fit of the model ($\chi^2_{MErr}=2130.70$, $p \leq 0.001$); keyword recognition declined significantly for sentences with masker errors ($\beta_{MErr}=-2.63$; $z=-27.62$; $p \leq 0.001$). However, masker errors did not interact with any other factor in the model, including age group and probe shift factors ($\chi^2<0.95$, *ns* in all cases)

Additional item-level and subject-level factors were also tested for significance. Sentence length and keyword position significantly improved the fit of the model (χ^2_{nWords}=121.41, $p \leq 0.001$; χ^2_{Pos}=89.01, $p \leq 0.001$); keyword recognition was significantly poorer for sentences with more keywords (β_{nWords}=-0.12; z=-11.01; $p \leq 0.001$) and improved significantly over the course of a sentence (β_{Pos}=0.35; z=17.93; $p \leq 0.001$). Subject-level differences in Connections and TRT were predictive of performance ($\chi^2_{Connections}$=9.64, $p \leq 0.01$; χ^2_{TRT}=10.20, $p \leq 0.01$); better keyword recognition was observed among subjects with faster speed of processing ($\beta_{Connections}$=0.19; z=3.29; $p \leq 0.001$) and better linguistic closure (β_{TRT}=0.20; z=3.41; $p \leq 0.001$). These results are consistent with findings from Experiments 1 and 2 and further support the contributions of these item-level and subject-level factors.

Figure 13 recasts the same data to show differences in keyword recognition and masker errors for sentences with a male target talker (left) and a female target talker (right). Model testing confirmed significant improvements in the fit of the model with the addition of talker sex (χ^2_{Sex}=15.42, $p \leq 0.001$), the interaction between talker sex and age ($\chi^2_{Sex*Age}$=21.56, $p \leq 0.001$), the interaction between talker sex and the 0.5 semitone probe shift ($\chi^2_{Probe0.5*Sex}$=25.13, $p \leq 0.001$), and the interaction between talker sex and the 0 semitone probe shift ($\chi^2_{Probe0*Sex}$=9.01, $p \leq 0.01$).

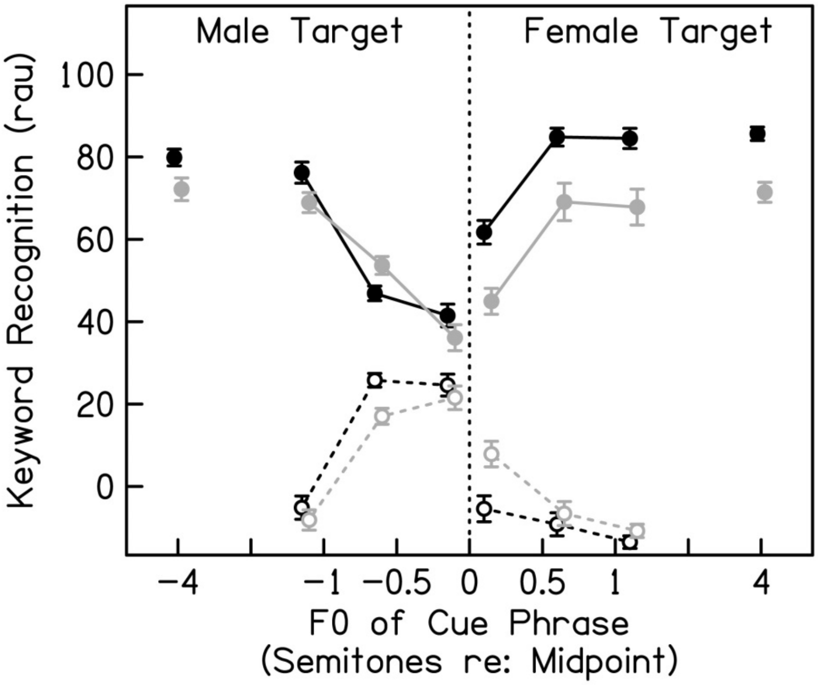

Figure 13: The same data from Figure 12 is recast to show differences between male target talkers (left) and female target talkers (right). Keyword recognition plotted in rau (solid lines) and masker errors (dotted lines) for younger (black) and older (gray) adults plotted as a function of semitone shift of the cue phrase. Probe trials with male target talker demonstrate poorer performance and more masker errors than trials with female target talker, particularly for 0.5 semitone shift. Large consistent effects of age are observed for recognition of female talkers for standard and probe trials.

Modeling results revealed that keyword recognition was better for sentences with a female talker than a male talker (β_{Sex}=0.15; z=3.96; $p \leq 0.001$). The interaction between age group and talker sex ($\beta_{Sex*Age}$=-0.10; z=-5.32; $p \leq 0.001$) was explored with a *post-hoc* model with separate talker sex factors for younger and older adults; results indicated that younger adults recognized significantly more keywords from female talkers than male talkers (β_{Sex_Y}=0.25;

z=5.95; $p \leq 0.001$), whereas differences in talker sex did not significantly predict keyword recognition for older adults (β_{Sex_O}=0.04; z=1.02; *ns*). The interactions between talker sex and the 0.5 semitone probe shift ($\beta_{Probe0.5*Sex}$=0.12; z=5.00; $p \leq 0.001$) and 0-semitone shift ($\beta_{Probe0*Sex}$=-0.07; z=-3.00; $p \leq 0.01$) were explored with *post-hoc* models with separate probe shift factors for male and female talkers. Results for the 0.5-semitone shift indicated that keyword recognition declined significantly for male talkers ($\beta_{Probe0.5\_M}$=-0.12; z=-3.78; $p \leq 0.001$) but not for female talkers ($\beta_{Probe0.5\_F}$=0.01; z=0.22; *ns*). Results for the 0-semitone shift indicated significant declines in keyword recognition for both male and female talkers, but declines were greater for female talkers (β_{Probe0_F}=-0.52; z=-18.21; $p \leq 0.001$) than for male talkers (β_{Probe0_M}=-0.27; z=-8.22; $p \leq 0.001$). Additional factors were tested for significance, including the four measures of hearing sensitivity described in Experiments 1 and 2, but model testing revealed that none of these additional factors significantly improved model fit ($\chi^2<3.07$, *ns* in all cases).

IV.D. Discussion

This study investigated the effects of age and listener expectations on object selection using a speech recognition task with a competing talker. On each trial, F0 and spectral envelope of the target and competing talker were adjusted for a standard difference corresponding to an 8 semitone difference in F0. This ensured that the perceptual cost of speech segregation was equivalent across trials. In order to assess the contribution of attention-based object selection, listener expectations of the target talker’s voice characteristics were manipulated with a cue phrase that preceded each trial. On the majority of trials (standard trials), the cue phrase matched the target's voice characteristics, which facilitated accurate selection of the target talker in the two-talker mixture. On a small proportion of randomly occurring probe trials, the voice characteristics of the cue phrase were parametrically shifted towards the midpoint between the target and competing talker. Larger shifts in the voice characteristics of the cue phrase resulted in poorer keyword recognition and a greater number of masker errors, whereby listeners responded with words from the competing talker sentence. These results suggest that listeners used the cue phrase to prime their attention for selection of an expected voice, resulting in enhanced recognition on standard trials relative to probe trials.

This application of the probe-signal method is similar to established methods used to study the role of attention in speech recognition (i.e., Humes et al., 2006; Kidd et al., 2005), but allows for assessment of attentional control in a more natural and dynamic way. At least two listening strategies may have been used on this task; (1) a "selective attention" strategy, in which a single talker was quickly selected as the focus of attention and the other talker was ignored, or (2) a "divided attention" strategy, in which attention was split between the two talkers until the listener was certain of the target talker. The selective attention strategy would have been highly successful on standard trials, as the cue phrase allowed for accurate selection of the target talker, but would have resulted in a large number of masker errors on probe trials. The divided attention strategy may have been more successful than selective attention on probe trials, but would have resulted in an unnecessary expenditure of cognitive resources on standard trials. As probe trials occurred randomly and listeners were not informed of the existence of probe trials, a divided attention strategy would have been impractical for all trials. Rather, listeners were expected to adopt a selective attention strategy for the majority of trials and deviate from this strategy only when it became clear that the cue phrase did not match either talker in the mixture. As such, the experimental design used in Experiment 3 provided a means to assess attentional control in a fluid and realistic manner, where listeners shift between selective and divided attention to maximize performance. Moreover, the design more closely reflects real-world communication practices where listeners expect to hear certain voices in a conversation, but occasionally, an unexpected voice is added that may or may not require the listener's attention.

IV.D.1. Attentional "tuning" for speech

The pattern of decline in performance on probe trials suggests fairly broad tuning of attention for voice features. Probe trials in which the voice characteristics of the cue phrase were shifted to 1 semitone from the midpoint between talkers resulted in essentially equivalent performance to standard trials. Considering the voice characteristics of the target talker were 4 semitones from the midpoint, this finding indicates that listeners were unaffected by deviations as large as 3 semitones from the expected voice characteristics based on the cue phrase. Interestingly, performance was above chance for probe trials in which the cue phrase was at the midpoint between talkers (i.e., 0-semitone shift). An intuitive hypothesis would propose that when the cue phrase was positioned exactly between the target and competing talker, listeners would be forced to select a talker at random, resulting in equal numbers of target keywords correct and masker errors. However, responses from both younger and older listeners for 0 semitone probe trials contained more correct keywords than masker errors, suggesting both groups were selecting the correct target talker with good accuracy. A likely explanation for this effect is that listeners may have used other cues to identify the target talker when F0 and spectral envelope were ambiguous. For example, the TIMIT corpus contains talkers from many different dialect regions. Because cue phrases were always taken from the target talker's recording of a standard sentence, listeners may have been able to identify the target talker based on dialectal variations, prosody, intonation, and other suprasegmental cues present in both the cue phrase and the target sentence. Thus, the broad attentional tuning for voice features in this study may reflect the multidimensionality of attentional tuning observed with non-speech sounds using probe-signal method (e.g., Dai & White, 1995; White & Carlyon, 1997; Wright & Dai, 1994).

IV.D.2. Effects of age on object selection

Hypotheses for Experiment 3 predicted that age-related declines in object selection would result in poorer performance by older adults compared to younger adults on standard trials, where focused attention on the voice characteristics of the cue phrase would be most beneficial to performance. On probe trials, the benefit of focused attention on the voice characteristics of the cue phrase was reduced, as these voice characteristics did not provide an accurate and efficient method of selecting the target talker. Performance of younger and older adults was expected to converge on probe trials as the cue phrase was shifted further towards the midpoint between the target and competing talker. This hypothesis was not supported by the data. Whereas older adults performed poorer than younger adults overall, no significant interaction was observed between age and cue phrase shift on probe trials. Several possible explanations may be considered for the persistent effect of age on probe trials. Younger adults are known to perform better than older adults on tasks that require divided attention (Humes et al., 2006). As such, younger adults may have been more adept at adjusting their listening strategy to facilitate performance on probe trials. Better performance was also predicted by faster cognitive speed of processing. Thus, age-related declines in speed of processing (e.g., Salthouse, 2000) may have limited the extent to which older adults could quickly and efficiently attend to an unexpected voice on probe trials. In contrast, variance in working memory capacity did not predict performance, suggesting that overall processing capacity was less critical for performance than fast and efficient use of available cognitive resources.

IV.D.3. Effects of talker sex

Differential effects of age were observed for recognition of sentences with male and female target talkers. As observed in Experiments 1 and 2, effects of age were more pronounced for recognition of female talkers than male talkers. The consistency of this effect across studies reveals that the age-related decline in recognition of female talkers is not specific to interrupted speech, but may reflect a broader outcome that includes the more typical continuous sentences used in Experiment 3. Furthermore, the effect of age on recognition of female talkers in this study was consistent across standard and probe trials, which suggests that age-related declines in recognition of female talkers is not uniquely dependent on object selection or the ability to focus attention on female voice characteristics. Rather, older adults may be less able to segregate a female target talker from a male competing talker. This interpretation explains the consistent size of the effect across trials, as the difference in voice features was 8 semitones from both standard and probe trials. Older adults may be less able to use periodicity cues present in the harmonic structure of a female talker to facilitate segregation and spectral glimpsing (Deroche et al., 2014). Larger spacing of harmonics may also lead to difficulty identifying formant position in female voices (Dissard & Darwin, 2001). These possibilities warrant further investigation of talker sex effects with greater controls for distinguishing effects related to F0 and harmonic structure.

The pattern of decline in recognition with the shift in the voice characteristics of the cue phrase was asymmetrical for male and female target talkers. Performance declined significantly for probe trials with a male target talker when the cue phrase was shifted to 0.5 semitones from the midpoint, whereas a similar shift value applied to a female cue phrase had no effect on performance. Declines in recognition of a female talker required a larger shift in the voice characteristics of the cue phrase to the midpoint between the target and competing talker (i.e., 0 semitone shift). This asymmetry may have been a consequence of the processing scheme used to shift the voice features of the cue phrase. More specifically, the geometric mean F0 of the two talkers was used for each individual talker pair as the reference (i.e., the midpoint) for subsequent processing of that sentence pair. This midpoint is likely different than the true perceptual midpoint between a “male-sounding” and “female-sounding” talker. F0 varies over a larger range among female talkers than male talkers (Peterson & Barney, 1952) and the average F0 of cue phrases in the 0 semitone probe condition may fall more closely into the range of a low-pitched female voice than a high-pitched male voice. This interpretation is supported by the visually apparent asymmetry in masker errors; probe trials with a male target talker yielded more masker errors than probes with a female target talker. Thus, when listeners’ attention was directed towards voice characteristics that were 0.5 or 0 semitones from the midpoint between talkers, they were more likely to respond with keywords from the female voice. This asymmetry is noteworthy but does not alter the basic findings of the experiment; that is, younger and older adults benefited from focused attention to expected voice features, unexpected deviations in the target talker’s voice resulted in a decline in speech recognition for both younger and older adults, and age-related differences in speech recognition were larger for female talkers than male talkers.

IV.E. Conclusions

A variation of the probe-signal method was used to explore effects of attention and listener expectations on object selection in a multi-talker environment. Overall, this method allowed for a natural assessment of listeners’ ability to use both selective and divided attention in a fluid manner to facilitate speech recognition in realistic listening conditions. Results indicated that recognition was poorer when the target's F0 and spectral envelope deviated from their expectations, suggesting that both younger and older listeners used focused attention and expectations of voice characteristics to facilitate object selection. Older adults performed poorer than younger adults overall, particularly for sentences with a female target talker, but both groups were equally affected by deviations in voice characteristics. This finding did not support the hypothesis that age-related declines in object selection contribute to speech recognition difficulty in older adults.

V. Conclusions

The primary goal of the three experiments in this research project was to determine the extent to which age-related declines in speech recognition in realistic listening environments could be explained by difficulties with specific components of the perceptual organization of speech. The framework presented in the Literature Review describes perceptual organization as the combined effects of several lower level auditory processes that facilitate formation of a coherent internal representation of speech (object formation) and higher level processes that listeners use to guide their attention to a talker of interest (object selection). Object formation in realistic listening environments is likely facilitated by temporal cues in speech corresponding to the slow modulations that code syllabic and segmental rates (envelope cues) and faster fluctuations that code F0 and intonation (periodicity cues). These cues are thought to provide separate contributions to three components of object formation: (1) “glimpsing,” the ability to connecting short segments of speech across time, (2) “phonemic restoration,” the ability to fill in missing information based on the available acoustic information, knowledge of language, and semantic context, and (3) “speech segregation,” the ability to perceptually separate glimpses of target speech from other competing speech sources. Object selection in realistic listening environments is likely guided by the listener’s intentions and expectations of a target talker’s voice. A listener’s expectations of the target talker may prime their attention to organize the auditory scene with the expected auditory object in the foreground and competing talkers in the background. Difficulty with any aspect of perceptual organization will result in a decline in speech recognition with a competing talker. As such, each of these component processes was viewed as a potential contributor to age-related declines in speech recognition in realistic listening environments.

A review of extant literature led to several hypotheses regarding the effects of age and the roles of envelope and periodicity cues on various components of perceptual organization. Envelope cues were hypothesized to improve glimpsing and facilitate phonemic restoration by providing a structured pattern for tracking the sentence through the momentary interruptions. Periodicity cues were expected to benefit speech segregation, by providing a continuous F0 trajectory to connect glimpses of target speech across time and avoid intrusions by competing speech. When both envelope and periodicity cues were available, their combined benefits were expected to be additive. Older adults were hypothesized to demonstrate poorer glimpsing than younger adults. Age-related declines in glimpsing were expected to be partially offset by greater phonemic restoration among older adults compared to younger adults. Older adults were hypothesized to benefit less than younger adults from periodicity cues, due to age-related declines in periodicity coding. Age-related declines in the use of periodicity cues were expected to result in poorer speech segregation for older adults compared to younger adults. Finally, age-related declines in attention were hypothesized to result in poorer speech recognition for older adults than younger adults when expected voice features would facilitate object selection. Hypotheses related to object formation and object selection were tested in separate experiments in order to avoid confounds between these two related processes.

Three experiments were designed to test these hypotheses. Experiments 1 and 2 evaluated age-related changes in the use of envelope and periodicity cues for glimpsing, phonemic restoration, and speech segregation. Experiment 3 evaluated the effects of age on the ability to focus attention on expected voice characteristics to facilitate selection of a target talker with an opposite sex competing talker. All subjects completed a battery of cognitive tasks to assess the contributions of processing speed, working memory capacity, inhibitory control, and linguistic closure to speech recognition performance and specific aspects of perceptual organization. Across all three experiments, data were analyzed with a logistic regression based on a GLMM. Several item-level factors, such as lexical characteristics of keywords, sentence length, keyword position, and talker sex, and several subject-level factors, including cognitive abilities, education level, and hearing sensitivity, were added to the model to test hypotheses regarding the contributions of these factors to speech recognition and perceptual organization. GLMM standard estimates for specific factors and interaction terms were used to evaluate hypotheses regarding the effects of age and temporal cues on speech recognition and perceptual organization. Several unexpected findings emerged as well, which indicated future directions for research. Overall, this work represents a substantial contribution to the growing body of literature on aging and perceptual organization of speech. Conclusions drawn from this work provide further insight into the effects of age on speech recognition in realistic listening environments.

V.A. Contributions of Envelope and Periodicity Cues

In Experiments 1 and 2, the relative contributions of envelope and periodicity cues to object formation were evaluated using sentences interrupted with silence, envelope cues, periodicity cues, or both cues. Results were in general agreement with recent work by Oh and colleagues (2016). In their study, recognition of interrupted vocoded speech was measured for conditions in which a continuous source of envelope cues, periodicity cues, or combined envelope and periodicity cues were presented to the contralateral ear. Results indicated roughly equivalent benefits associated with contralateral envelope or periodicity cues, and these benefits were additive when both cues were presented to the contralateral ear. Their interpretation of the results was similar to those proposed in Experiments 1 and 2; that is, continuity of envelope and/or periodicity cues over the course of the sentence facilitated integration of speech glimpses into a single coherent representation of the sentence. In Experiments 1 and 2, envelope and periodicity cues were presented ipsilaterally, rather than contralaterally, but still provided a continuous source of envelope and/or periodicity cues when combined with glimpses of speech. In both studies, these cues facilitated object formation, indicating that the mechanism that incorporates glimpses of speech into a single coherent object receives both ipsilateral and contralateral projections.

In Experiment 2, periodicity cues were expected to benefit speech segregation by providing a continuous representation of the target talker's F0. This hypothesis was not supported by the data. A recent study by Steinmetzger and Rosen (2015) also indicated limited contributions of target periodicity to speech recognition. In their study, sentence stimuli were either unprocessed or processed through one of three vocoders designed to retain different amounts of periodicity information. These sentences were presented in a background of aperiodic noise or pulse trains that carried a time-varying F0. Their results indicated that speech recognition depended minimally on the amount of periodicity information retained in the target. Rather, the largest improvements in performance were obtained when the masker contained periodicity cues. Benefits of masker periodicity were observed for steady-state maskers as well as 10 Hz modulated maskers, which allowed for glimpsing of target speech. They interpreted these results as evidence for harmonic cancellation theory (de Cheveigné, McAdams, Laroche, & Rosenberg, 1995; de Cheveigné, McAdams, & Marin, 1997), which states that masker periodicity allows the auditory system to effectively subtract the harmonically related masker energy from the signal. In Experiments 1 and 2, the masker was always an unprocessed competing talker and thus contained a consistent amount of periodicity information across the different sentence types. As such, Experiments 1 and 2 are not well suited to determine the extent to which harmonic cancellation theory may account for the benefit associated with masker periodicity. However, periodicity cues improved recognition both in quiet and with a competing talker, indicating that restoring continuity of pitch and intonation through interrupted segments provided a benefit to speech recognition.

V.B. Effects of Age on Object Formation

Age-related declines in object formation were hypothesized to partially explain speech recognition difficulty of older adults in realistic listening environments. Results of Experiments 1 and 2 indicated considerable age-related declines in glimpsing, which were most pronounced when speech glimpses were short. When envelope and periodicity cues were available, older adults were able to use these supportive cues to facilitate glimpsing and restoration of missing speech information. However, speech recognition by older adults did not improve to the level of younger adults. Rather, residual age-related difficulty was observed even when both envelope and periodicity cues were available. Thus, any temporal disruptions to the speech signal had a more detrimental effect on speech recognition for older adults than younger adults. These results are in agreement with established theories of age-related declines in temporal processing and its effects on speech recognition (Fitzgibbons & Gordon-Salant, 1995; 1996, Gordon-Salant and Fitzgibbons, 1993). One consequence of the age-related decline in temporal processing is poorer resolution of silent temporal intervals (Fitzgibbons & Gordon-Salant, 1987; 1994; Gordon-Salant et al., 2006), which may considerably disrupt recognition of silence-interrupted speech. Another consequence of declines in temporal processing is poor resolution of low-frequency periodicity cues in speech (Pichora-Fuller, 2007). This may explain why older adults were less able to use the available glimpses of speech than younger adults, even when additional temporal cues were available to facilitate glimpsing. Thus, the age-related decline in glimpsing observed here is consistent with existing theories regarding the effects of poor temporal resolution on speech recognition in older adults.

The addition of a competing talker resulted in poorer performance for both younger and older adults, consistent with a perceptual cost associated with speech segregation. This cost was hypothesized to be greater for older adults, based on existing theories of the effects of age on speech recognition in backgrounds with competing speech (Helfer & Freyman, 2008; Lee & Humes, 2012; Rajan & Cainer, 2008). However, any effects of age on speech segregation had only a minor effect on performance. One possible reason for this deviation from existing work on aging and competing speech relates to the use of interrupted speech in these studies. Glimpsing segments of speech is a critical component of this task and serves as a primary predictor of performance. The additional demands associated with speech segregation may contribute less to the overall variance in performance on this task than glimpsing. As such, age-related declines in glimpsing may outweigh effects of age on speech segregation observed in studies with continuous speech (Helfer & Freyman, 2008; Lee & Humes, 2012; Rajan & Cainer, 2008). Considering that large effects of age were observed in Experiments 1 and 2 even in quiet, these results support the hypothesis that the primary effect of aging on object formation is a decline in the ability to connect glimpses of speech across time.

V.C. Speech Segregation vs. Target Selection

One goal of Experiment 3 was to dissociate effects of speech segregation from target selection. This was achieved by setting a fixed difference in voice features between the target and competing talker sentence on each trial. Object selection was manipulated separately based on the voice characteristics of the cue phrase. As such, effect of focused attention and object selection could be evaluated without confounds associated with additional speech segregation cues. Though the results did not support the hypothesis regarding age-related declines in object selection, the method was successful in demonstrating a decline in recognition associated with unexpected voice features. This method provides some advantages over existing designs comparing selective and divided attention on separate blocks of trials (i.e., Lee & Humes, 2012). Using separate trials allows listeners to develop specific task sets that are not representative of real-world applications of attention. By evaluating selective and divided attention in a more fluid manner across trials, the dynamic ability to shift attention and accommodate unexpected changes in stimuli can be assessed. Future studies using this method should explore other dimensions that listeners may use for talker selection, such as prosodic or dialectal cues. The TIMIT corpus is well suited for such an investigation, as talkers are organized by dialectal region.

V.D. Effects of Age and Talker Sex

A persistent finding across all three experiments was an age-related decline in recognition of female talkers. This finding was independent of variance in hearing sensitivity or the presence of supportive periodicity cues. The sentence material used in these studies contained a diverse sample of male and female talkers, who were equally intelligible to younger adults in quiet. Results of Experiment 3 indicated that the effects of age on talker sex were independent of the locus of attention; effects of age on recognition of female talkers were consistent across standard and probe trials. There are several possible explanations for the observed decline in recognition of female voices in older adults. Age-related declines in periodicity coding and resolution of harmonic structure in older adults may disproportionately affect the recognition of female talkers (Clinard & Tremblay, 2013; Souza et al., 2011). Broader spacing of harmonic structure may also disrupt recognition of formant frequencies that occur between harmonic peaks (Dissard & Darwin, 2001). Female talkers also typically have more dynamic pitch contours. Whereas dynamic pitch may provide an advantage in younger adults, older adults may be less able to benefit from these dynamic temporal cues (Shen, Wright & Souza, 2016; Souza et al., 2011). The precise combination of factors that contribute to this age-related decline in recognition of female talkers remains unclear and warrants further study.

V.E. Role of Cognitive Abilities in Speech Recognition

Cognitive test results across Experiments 1, 2, and 3 indicated that two measures provided the best predictions of keyword recognition, Connections and TRT. The TRT has previously been found to predict recognition of interrupted speech (Krull et al., 2013). This measure provides an intuitive visual parallel and suggests an amodal ability to use partial linguistic information may contribute to recognition of interrupted speech. The Connections test is a measure of speed of processing, which reflects the efficiency with which participants can manipulate information. As speech naturally unfolds over time, the ability to recognize glimpsed speech and restore missing information requires quick and efficient cognitive processing. This task is different from many other speech recognition tasks, such as measures of a speech recognition threshold (SRT). SRT and other speech measures have been found to correlate with working memory capacity in several studies (Foo et al., 2007; Gordon-Salant & Cole, 2016; Lunner, 2003; Rudner et al., 2009; Souza & Arehart, 2015). In contrast, no relationship was observed in Experiments 1, 2, or 3 between speech recognition and working memory capacity. The use of a visual test of working memory capacity may have limited the strength of the relationship between the cognitive measure and speech recognition. Future studies should explore this relationship with an auditory based measure of working memory, such as the Listening Span (e.g., Gordon-Salant & Cole, 2016). These results suggest that different speech recognition tasks may place different demands on distinct cognitive abilities. Future studies should carefully consider the cognitive demands associated with the speech recognition task in order to select the most appropriate cognitive measures.

V.F. Limitations

The use of silence-interrupted sentences as the baseline condition limits the extent to which the results of Experiments 1 and 2 can characterize the relative contributions of envelope and periodicity cues. The presence of silent gaps likely disrupted performance in the baseline condition, particularly for older adults. This disruption is not characteristic of a lack of temporal cues, but rather the presence of erroneous silent intervals in an otherwise continuous sentence. As such, the current results may overestimate the contributions of envelope and periodicity cues, as the filled sentences eliminated the silent gaps in addition to providing additional speech cues. Recognition of steady-state-noise-filled sentences would have provided a better baseline condition that contained neither envelope nor periodicity and was also free of silent gaps. Improvements in recognition associated with modulating the noise by the missing envelope or replacing the noise with a periodicity carrier would have provided a clearer picture of the true contributions of these temporal cues to glimpsing. However, the limited number of PRESTO lists precluded the addition of a steady-state-noise-filled condition, which would have served primarily as a replication of previous work (i.e., Bashford et al., 1996; Shinn-Cunningham & Wang, 2008).

Speech segregation results in these studies are limited to conditions with a single opposite-sex competing talker. As such, there were always considerable acoustic differences between target and competing talkers in F0 and spectral envelope, which likely facilitated speech segregation. In addition, a single competing talker provides deeper modulations to facilitate glimpsing relative to speech babble used in other studies (Ben-David et al., 2012; Rajan & Cainer, 2008). This may have contributed to the relatively minor effects of age observed in these studies relative to existing literature on aging and speech recognition with competing talkers. Pairing of opposite-sex talkers also limited the extent to which the effects of talker sex could be attributed to a male/female target, rather than a male/female competing talker. Use of same-sex talker pairs would have provided useful data to determine whether the age-related decline in recognition of female talkers depended on the sex of the competing talker. However, recognition of interrupted speech is a difficult task, particularly with a competing talker, and use of multi-talker babble or same-sex talker pairs would have likely resulted in additional floor effects.

The limited number of sentences available in the PRESTO corpus necessitated repetition of sentences across experiments. Repeated exposure to the PRESTO sentence lists resulted in improved performance for both younger and older adults across Experiments 1 and 2. It was not possible to determine what component of this learning effect was related to prior experience with PRESTO sentences, rather than more general task learning associated with interrupted speech. In addition, prior experience with the sentence material likely enhanced performance of all subjects in Experiment 3, which was always completed last. These learning effects were quantified as well as possible within the GLMM, such that this additional source of variance could be accounted for within the model.

V.G. Conclusion

The three experiments in this research project revealed that age-related declines in speech recognition can be partially explained by declines in several components of perceptual organization. The most pronounced effects of age were on the ability to connect short glimpses of speech across time. These results demonstrate functional declines in speech recognition that likely stem from known deficits in temporal resolution in older adults. Additional unexpected findings emerged including an age-related decline in recognition of female talkers. This work adds to the growing body of literature on aging and speech recognition and strengthens connections between effects of aging and perceptual organization of speech. The methods used here can be adapted to address new research question that follow from these results. Future work will be designed to address these new research questions.

Appendix 1: Experiment 1 GLMM standard estimates, standard error, and *z*-statistics are displayed for each significant fixed effect and interaction term in the order they appear in the text. Each Interaction was explored with a separate post-hoc model with split factors, and statistical results are indented below interactions. Asterisks indicated significance levels for *z*-statistics (*** < 0.001; ** < 0.01; * < 0.05).

Factor	**Standard Estimate (β)**	**Standard Error**	***z*-Statistic**
Age Group (*Age*)	-0.29	0.06	-5.23***
Proportion (*Prop*)	1.03	0.02	58.41***
Interruption Type (*Int*)	0.23	0.02	14.51***
Background (*Bg*)	-0.75	0.02	-39.11***
Proportion × Background (_Prop*Bg_)	0.10	0.01	6.62***
Proportion, Competing Talker (*Prop_CT*)	1.12	0.03	38.57***
Proportion, Quiet (*Prop_Q*)	0.93	0.02	42.33***
Proportion × Age Group (_Prop*Age_)	0.07	0.02	4.43***
Proportion, Older (*Prop_O*)	1.06	0.02	42.79***
Proportion, Younger (*Prop_Y*)	0.92	0.02	39.13***
Proportion × Interruption Type (_Prop*Int_)	-0.07	0.01	-5.42***
Proportion, Silence Interrupted (*Prop_SInt*)	1.06	0.03	42.14***
Proportion, Envelope Filled (*Prop_Env*)	0.92	0.02	38.44***
Interruption Type × Age Group (_Int*Age_)	0.06	0.02	3.61***
Interruption Type, Older (*Int_O*)	0.26	0.02	11.95***
Interruption Type, Younger (*Int_Y*)	0.16	0.02	7.59***
Background × Interruption Type (_Bg*Int_)	0.03	0.01	3.03**
Background, Silence Interrupted (*Bg_SInt*)	-0.73	0.02	-34.79***
Background, Envelope Filled (*Bg_Env*)	-0.68	0.02	-33.58***
Word Frequency (*WF*)	0.28	0.01	27.72***
Biphone Probability (*BP*)	0.03	0.01	2.90**
Neighborhood Density (*ND*)	-0.15	0.01	-14.00***

Appendix 1 continued…

Factor	**Standard Estimate (β)**	**Standard Error**	***z*-Statistic**
Word Frequency × Interruption Type (*Int*WF*)	-0.02	0.01	-2.32*
Interruption Type, Uncommon Words (*Int_loWF*)	0.25	0.01	18.64***
Interruption Type, Common Words (*Int_hiWF*)	0.20	0.02	11.28***
Sentence Length (*nWords*)	-0.13	0.01	-12.45***
Keyword Position (*Pos*)	0.17	0.01	16.38***
Keyword Position × Age Group (*Pos*Age*)	-0.04	0.01	-3.75***
Keyword Position, Younger (*Pos_Y*)	0.19	0.01	13.64***
Keyword Position, Older (*Pos_O*)	0.11	0.01	7.09***
Sentence Length × Background (*nWords*Bg*)	-0.09	0.01	-7.99***
Sentence Length, Competing Talker (*nWords_CT*)	-0.22	0.02	-12.45***
Sentence Length, Quiet (*nWords_Q*)	-0.05	0.01	-3.89***
Keyword Position × Background (*Pos*Bg*)	0.11	0.01	10.12***
Keyword Position, Competing Talker (*Pos_CT*)	0.28	0.02	17.10***
Keyword Position, Quiet (*Pos_Q*)	0.07	0.01	5.28***
Talker Sex × Age Group (*Sex*Age*)	-0.08	0.01	-7.95***
Talker Sex, Younger (*Sex_Y*)	0.07	0.01	5.29***
Talker Sex, Older (*Sex_O*)	-0.07	0.02	-4.70***
Talker Sex × Background (*Sex*Bg*)	0.03	0.01	3.03**
Talker Sex, Quiet (*Sex_Q*)	-0.03	0.01	-2.54*
Talker Sex, Competing Talker (*Sex_CT*)	0.04	0.02	2.18*
Talker Sex × Background × Age Group (*Sex*Bg*Age*)	-0.04	0.01	-4.00***
Talker Sex, Quiet, Younger (*Sex_Q_Y*)	0.01	0.02	0.53*ns*
Talker Sex, Quiet, Older (*Sex_Q_O*)	-0.07	0.02	-3.91***
Talker Sex, Competing Talker, Younger (*Sex_CT_Y*)	0.15	0.02	7.02***
Talker Sex, Competing Talker, Older (*Sex_CT_O*)	-0.09	0.02	-3.84***
Education Level (*Edu*)	0.10	0.04	2.35*

Appendix 1 continued…

Factor	**Standard Estimate (β)**	**Standard Error**	***z*-Statistic**
Connections (*Connections)*	0.13	0.05	2.80**
TRT (*TRT*)	0.13	0.05	2.60**
TRT × Interruption Type (*TRT*Int*)	-0.06	0.01	-4.19***
TRT, Silence Interrupted (*TRT_SInt*)	0.19	0.05	3.40***
TRT, Envelope Filled (*TRT_Env*)	0.07	0.05	1.44*ns*
Interruption Type, Poorer TRT (*Int_loTRT*)	0.27	0.02	12.40***
Interruption Type, Better TRT (*Int_hiTRT*)	0.21	0.02	9.53***

Appendix 2: Experiment 2 GLMM standard estimates, standard error, and *z*-statistics are displayed for each significant fixed effect and interaction term in the order they appear in the text. Each interaction was explored with a separate post-hoc model with split factors, and statistical results are indented below interactions. Asterisks indicated significance levels for *z*-statistics (*** < 0.001; ** < 0.01; * < 0.05).

Factor	**Standard Estimate (β)**	**Standard Error**	***z*-Statistic**
Experiment Order (*Ord*)	0.05	0.01	3.92***
Age Group (*Age*)	-0.27	0.05	-5.42***
Proportion (*Prop*)	1.00	0.01	77.60***
Background (*Bg*)	-0.75	0.01	-53.05***
Envelope Cues (*Env*)	0.24	0.01	20.82***
Periodicity Cues (*Prd*)	0.16	0.01	10.81***
Proportion × Background (*Prop*Bg*)	0.11	0.01	11.45***
Proportion, Competing Talker (*Prop_CT*)	1.10	0.02	65.88***
Proportion, Quiet (*Prop_Q*)	0.88	0.01	82.28***
Proportion × Age Group (*Prop*Age*)	0.06	0.01	4.91***
Proportion, Older (*Prop_O*)	1.00	0.01	75.48***
Proportion, Younger (*Prop_Y*)	0.89	0.01	71.90***
Background × Age Group (*Bg*Age*)	-0.04	0.01	-2.78**
Background, Older (*Bg_O*)	-0.74	0.01	-67.58***
Background, Younger (*Bg_Y*)	-0.68	0.01	-66.99***
Proportion × Envelope Cues (*Prop*Env*)	-0.10	0.01	-9.80***
Proportion, Envelope Off (*Prop_xEnvx*)	1.03	0.01	77.47***
Proportion, Envelope On (*Prop_Env*)	0.87	0.01	69.96***
Proportion × Periodicity Cues (*Prop*Prd*)	-0.04	0.01	-4.45***
Proportion, Periodicity Off (*Prop_xPrdx*)	0.98	0.01	76.89***
Proportion, Periodicity On (*Prop_Prd*)	0.91	0.01	74.18***

Appendix 2 continued…

Factor	**Standard Estimate (β)**	**Standard Error**	***z*-Statistic**
Envelope Cues × Age Group (*Env*Age*)	0.06	0.01	5.91***
Envelope Cues, Older (*Env_O*)	0.26	0.01	25.75***
Envelope Cues, Younger (*Env_Y*)	0.15	0.01	15.90***
Envelope Cues × Background (*Env*Bg*)	0.05	0.01	5.63***
Envelope Cues, Competing Talker (*Env_CT*)	0.22	0.01	19.75***
Envelope Cues, Quiet (*Env_Q*)	0.19	0.01	21.65***
Proportion × Envelope Cues × Age Group (*Prop*Env*Age*)	-0.03	0.01	-2.98**
Proportion × Envelope Cues, Older (*Prop*Env_O*)	-0.10	0.01	-7.61***
Proportion × Envelope Cues, Younger (*Prop*Env_Y*)	-0.05	0.01	-3.98***
Proportion × Periodicity Cues × Age Group (*Prop*Prd*Age*)	-0.02	0.01	-2.46*
Proportion × Periodicity Cues, Younger (*Prop*Prd_Y*)	-0.02	0.01	-1.38*ns*
Proportion × Periodicity Cues, Older (*Prop*Prd_O*)	-0.06	0.01	-4.86***
Proportion × Envelope Cues × Background (*Prop*Env*Bg*)	-0.03	0.01	-2.59**
Proportion × Envelope Cues, Competing Talker (*Prop*Env_CT*)	-0.11	0.02	-6.70***
Proportion × Envelope Cues, Quiet (*Prop*Env_Q*)	-0.06	0.01	-5.88***
Word Frequency (*WF*)	0.27	0.01	38.38***
Biphone Probability (*BP*)	0.04	0.01	6.13***
Neighborhood Density (*ND*)	-0.15	0.01	-20.32***
Envelope Cues × Word Frequency (*Env*WF*)	-0.02	0.01	-2.70**
Envelope Cues, Less Common Words (*Env_loWF*)	0.25	0.01	20.37***
Envelope Cues, More Common Words (*Env_hiWF*)	0.21	0.02	13.96***
Sentence Length (*nWords*)	-0.15	0.01	-20.24***
Sentence Length × Background (*nWords*Bg*)	-0.09	0.01	-12.43***
Sentence Length, Competing Talker (*nWords_CT*)	-0.24	0.01	-20.06***
Sentence Length, Quiet (*nWords_Q*)	-0.05	0.01	-6.52***

Appendix 2 continued…

Factor	**Standard Estimate (*β*)**	**Standard Error**	***z*-Statistic**
Keyword Position (*Pos*)	0.13	0.01	17.60***
Keyword Position × Age Group (*Pos*Age*)	-0.03	0.01	-4.38***
Keyword Position, Younger (*Pos_Y*)	0.14	0.01	14.79***
Keyword Position, Older (Pos_O)	0.07	0.01	7.38***
Keyword Position × Background (*Pos*Bg*)	0.07	0.01	10.14***
Keyword Position, Competing Talker (*Pos_CT*)	0.20	0.01	18.13***
Keyword Position, Quiet (*Pos_Q*)	0.05	0.01	5.67***
Keyword Position × Envelope Cues (*Pos*Env*)	0.01	0.01	2.14*
Keyword Position, Envelope On (*Pos_Env*)	0.13	0.01	13.15***
Keyword Position, Envelope Off (*Pos_xEnvx*)	0.09	0.01	9.14***
Keyword Position × Periodicity Cues (*Prd*Pos*)	-0.02	0.01	-2.71**
Keyword Position, Periodicity Off (*Pos_xPrdx*)	0.13	0.01	12.84***
Keyword Position, Periodicity On (*Pos_Prd*)	0.09	0.01	9.61***
Talker Sex × Age Group (*Sex*Age*)	-0.08	0.01	-11.23***
Talker Sex, Younger (*Sex_Y*)	0.08	0.01	8.66***
Talker Sex, Older (*Sex_O*)	-0.07	0.01	-6.99***
Talker Sex × Background (*Sex*Bg*)	0.04	0.01	5.99***
Talker Sex, Quiet (*Sex_Q*)	-0.04	0.01	-4.34***
Talker Sex, Competing Talker (*Sex_CT*)	0.05	0.01	4.51***
Talker Sex × Background × Age Group (*Sex*Bg*Age*)	-0.02	0.01	-2.47*
Talker Sex, Quiet, Younger (*Sex_Q_Y*)	0.03	0.01	2.06*
Talker Sex, Quiet, Older (*Sex_Q_O*)	-0.10	0.01	-8.01***
Talker Sex, Competing Talker, Younger (*Sex_CT_Y*)	0.15	0.01	9.86***
Talker Sex, Competing Talker, Older (*Sex_CT_O*)	-0.05	0.02	-3.08**
Education Level (*Edu*)	0.10	0.04	2.47*
Connections (*Connections)*	0.15	0.04	3.59***

Appendix 3: Experiment 3 GLMM standard estimates, standard error, and *z*-statistics are displayed for each significant fixed effect and interaction term in the order they appear in the text. Each interaction was explored with a separate post-hoc model with split factors, and statistical results are indented below interactions. Asterisks indicated significance levels for *z*-statistics (*** < 0.001; ** < 0.01; * < 0.05).

Factor	**Standard Estimate (*β*)**	**Standard Error**	***z*-Statistic**
Age Group (*Age*)	-0.24	0.06	-4.30***
Probe Shifted to 0.5 Semitones (*Probe0.5*)	-0.06	0.02	-2.28*
Probe Shifted to 0 Semitones (*Probe0*)	-0.40	0.02	-17.00***
Masker Errors (*MErr*)	-2.63	0.10	-27.62***
Sentence Length (*nWords*)	-0.12	0.01	-11.01***
Keyword Position (*Pos*)	0.35	0.02	17.93***
Connections (*Connections*)	0.19	0.06	3.29***
TRT (*TRT*)	0.20	0.06	3.41***
Talker Sex (*Sex*)	0.15	0.04	3.96***
Talker Sex × Age Group (*Sex*Age*)	-0.10	0.02	-5.32***
Talker Sex, Younger (*Sex_Y*)	0.25	0.04	5.95***
Talker Sex, Older (*Sex_O*)	0.04	0.04	1.02*ns*
Probe Shifted to 0.5 Semitones × Talker Sex (*Probe0.5*Sex*)	0.12	0.02	5.00***
Probe Shifted to 0.5 Semitones, Male (*Probe0.5_M*)	-0.12	0.03	-3.78***
Probe Shifted to 0.5 Semitones, Female (*Probe0.5_F*)	0.01	0.03	0.22*ns*
Probe Shifted to 0 Semitones × Talker Sex (*Probe0*Sex*)	-0.07	0.02	-3.00**
Probe Shifted to 0 Semitones, Female (*Probe0_F*)	-0.52	0.03	-18.21***
Probe Shifted to 0 Semitones, Male (*Probe0_M*)	-0.27	0.03	-8.22***

Literature Cited

Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *J Acoust Soc Am, 126*(4), 1926-1940.

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. *Int J Audiol, 47*(Suppl 2), S53-S71.

Alain, C., Ogawa, K. H., & Woods, D. L. (1996). Aging and the segregation of auditory stimulus sequences. *J Gerontol B Psychol Sci Soc Sci, 51*(2), P91-93.

ANSI. (2010). *American national standard specification for audiometers* (ANSI/ASA S3.6-2010), New York: American National Standards Institute.

Arehart, K. H., Souza, P. E., Muralimanohar, R. K., & Miller, C. W. (2011). Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing. *J Speech Lang Hear Res, 54*(1), 190-210.

Assmann, P. F., & Summerfield, Q. (2004). The perception of speech under adverse acoustic conditions. In S. Greenberg, W. A. Ainsworth, A. N. Popper, & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 231-308). New York: Springer-Verlag.

Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *J Speech Lang Hear Res, 41*(3), 549-563.

Bashford, J. A., Jr., Meyers, M. D., Brubaker, B. S., & Warren, R. M. (1988). Illusory continuity of interrupted speech: speech rate determines durational limits. *J Acoust Soc Am, 84*(5), 1635-1638.

Bashford, J. A., Jr., Riener, K. R., & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Percept Psychophys, 51*(3), 211-217.

Bashford, J. A., Jr., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Percept Psychophys, 42*(2), 114-121.

Bashford, J. A., Jr., Warren, R. M., & Brown, C. A. (1996). Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. *Percept Psychophys, 58*(3), 342-350.

Bașkent, D. (2010). Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. *J Acoust Soc Am, 128*(4), EL169-174.

Bașkent, D. (2012). Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. *J Assoc Res Otolaryngol, 13*(5), 683-692.

Bașkent, D., & Chatterjee, M. (2010). Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hear Res, 270*(1-2), 127-133.

Bașkent, D., Eiler, C., & Edwards, B. (2009). Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. *J Acoust Soc Am, 125*(6), 3995-4005.

Bașkent, D., Eiler, C. L., & Edwards, B. (2010). Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. *Hear Res, 260*(1-2), 54-62.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Ben-David, B. M., Tse, V. Y., & Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear Res, 290*(1-2), 55-63.

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proc Natl Acad Sci U S A, 105*(35), 13174-13178.

Bier, B., Lecavalier, N. C., Malenfant, D., Peretz, I., & Belleville, S. (2017). Effect of age on attentional control in dual-tasking. *Exp Aging Res, 43*(2), 161-177.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res, 27*, 32-48.

Boersma, P. & Weenick, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 24 November 2014 from http://www.praat.org/

Botte, M. C. (1995). Auditory attentional bandwidth: effect of level and frequency range. *J Acoust Soc Am, 98*(5 Pt 1), 2475-2485.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Mass.: MIT Press.

Bregman, A. S., Colantonio, C., & Ahad, P. A. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Percept Psychophys, 61*(2), 195-205.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *J Phon, 10*, 23-36.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am, 109*(3), 1101-1109.

Brysbaert, M. & New, B. (2009) Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods, 41*(4), 977-990.

Buus, S. (1985). Release from masking caused by envelope fluctuations. *J Acoust Soc Am, 78*(6), 1958-1965.

Chatterjee, M., Peredo, F., Nelson, D., & Başkent, D. (2010). Recognition of interrupted sentences under conditions of spectral degradation. *J Acoust Soc Am, 127*(2), EL37-41.

Chintanpalli, A., Ahlstrom, J. B., & Dubno, J. R. (2014). Computational model predictions of cues for concurrent vowel identification. *J Assoc Res Otolaryngol, 15*(5), 823-837.

Chintanpalli, A., Ahlstrom, J. B., & Dubno, J. R. (2016). Effects of age and hearing loss on concurrent vowel identification. *J Acoust Soc Am, 140*(6), 4142.

Clark, H. H. (1973). The Language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J Verbal Learning Verbal Behav, 12*(4), 335-359.

Clarke, J., Gaudrain, E., Chatterjee, M., & Başkent, D. (2014). T'ain't the way you say it, it's what you say--perceptual continuity of voice and top-down restoration of speech. *Hear Res, 315*, 80-87.

Clinard, C. G., & Tremblay, K. L. (2013). Aging degrades the neural encoding of simple and complex sounds in the human brainstem. *J Am Acad Audiol, 24*(7), 590-599.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *J Exp Psychol Hum Percept Perform, 16*(3), 551-563.

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: the use of context by children and adults. *J Verbal Learn Verbal Behav, 19*, 297-315.

Committee on Hearing, Bioacoustics, and Biomechanics (CHABA). (1988). Speech understanding and aging. *J Acoust Soc Am, 83*(3), 859-895.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *J Acoust Soc Am, 119*(3), 1562-1573.

Dai, H., & Wright, B. A. (1995). Detecting signals of unexpected or uncertain durations. *J Acoust Soc Am, 98*(2 Pt 1), 798-806.

Dai, H. P., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *J Acoust Soc Am, 89*(6), 2837-2842.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *J Verbal Learn Verbal Behav, 19*, 450-466.

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn Sci, 1*(9), 327-333.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am, 114*(5), 2913-2922.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 10*(2), 85-98.

de Cheveigné, A., McAdams, S., Laroche, J., & Rosenberg, M. (1995). Identification of concurrent harmonic and inharmonic vowels: a test of the theory of harmonic cancellation and enhancement, *J Acoust Soc Am, 97*(6), 3736-3748.

de Cheveigné, A., McAdams, S., & Marin, C. M. (1997). Concurrent vowel identification. II. Effects of phase, harmonicity, and task, *J Acoust Soc Am, 101*(5), 2848-2856.

Deroche, M. L., Culling, J. F., & Chatterjee, M. (2014). Phase effects in masking by harmonic complexes: detection of bands of speech-shaped noise. *J Acoust Soc Am, 136*(5), 2726-2736.

Deroche, M. L., Culling, J. F., Chatterjee, M., & Limb, C. J. (2014). Roles of target and masker fundamental frequencies in voice segregation. *J Acoust Soc Am, 136*(3), 1225-1236.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A, 109*(29), 11854-11859.

Dirks, D. D., Bell, T. S., Rossman, R. N., & Kincaid, G. E. (1986). Articulation index predictions of contextually dependent words. *J Acoust Soc Am, 80*(1), 82-92.

Dissard, P., & Darwin, C. J. (2001). Formant-frequency matching between sounds with different bandwidths and on different fundamental frequencies. *J Acoust Soc Am, 110*(1), 409-415.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am, 97*(1), 585-592.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *J Acoust Soc Am, 111*(6), 2897-2907.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: masking of speech by interrupted noise for younger and older adults with normal hearing. *J Acoust Soc Am, 113*(4 Pt 1), 2084-2094.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *J Exp Psychol Learn Mem Cogn, 20*(2), 470-484.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J Acoust Soc Am, 74*(3), 739-743.

Eisenberg, L. S., Dirks, D. D., & Bell, T. S. (1995). Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. *J Speech Hear Res, 38*(1), 222-233.

Ezzatian, P., Li, L., Pichora-Fuller, K., & Schneider, B. A. (2015). Delayed stream segregation in older adults: more than just informational masking. *Ear Hear, 36*(4), 482-484.

Faulkner, K. F., Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2015). List Equivalency of PRESTO for the Evaluation of Speech Recognition. *J Am Acad Audiol, 26*(6), 582-594.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am, 88*(4), 1725-1736.

Fitzgibbons, P. J., & Gordon-Salant, S. (1987). Temporal gap resolution in listeners with high-frequency sensorineural hearing loss. *J Acoust Soc Am, 81*(1), 133-137.

Fitzgibbons, P. J., & Gordon-Salant, S. (1994). Age effects on measures of auditory duration discrimination. *J Speech Hear Res, 37*(3), 662-670.

Fitzgibbons, P. J., & Gordon Salant, S. (1995). Age effects on duration discrimination with simple and complex stimuli. *J Acoust Soc Am, 98*(6). 3140-3145.

Fitzgibbons, P. J., & Gordon Salant, S. (1996). Auditory temporal processing in elderly listeners. *J Am Acad Audiol, 7*(3), 183-189.

Fitzgibbons, P. J., & Gordon Salant, S. (2001). Aging and temporal discrimination in auditory sequences. *J Acoust Soc Am, 109*(6), 2955-2963.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res, 12*(3), 189-198.

Foo, C., Rudner, M., R., Rönnberg, J. & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J Am Acad Audiol. 18*, 533-566.

Freyman, R. L., Griffin, A. M., & Oxenham, A. J. (2012). Intelligibility of whispered speech in stationary and modulated noise maskers. *J Acoust Soc Am, 132*(4), 2514-2523.

Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front Aging Neurosci, 6*, 347.

Füllgrabe, C., & Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Front Psychol, 7*, 1268.

Gallun, F. J., Mason, C. R., & Kidd, G. Jr. (2007). Task-dependent costs in processing two simultaneous auditory stimuli. *Percept Psychophys, 69*(5), 757-771.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1993). The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Philadelphia: Linguistic Data Consortium.

Gaudrain, E., & Carlyon, R. P. (2013). Using Zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation. *J Acoust Soc Am, 133*(1), 502-518.

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J. C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hear Res, 231*(1-2), 32-41.

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J. C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *J Acoust Soc Am, 124*(5), 3076-3087.

George, E. L. H., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M. & Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *J Acoust Soc Am, 121*, 2362-2375.

Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *J Acoust Soc Am, 110*(3), 1628-1640.

Gilbert, G., Bergeras, I., Voillery, D., & Lorenzi, C. (2007). Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. *J Acoust Soc Am, 122*(3), 1336.

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *J Am Acad Audiol, 24*(1), 26-36.

Gnansia, D., Pressnitzer, D., Pean, V., Meyer, B., & Lorenzi, C. (2010). Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. *Hear Res, 265*(1-2), 46-53.

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing, *Ear Hear, 37*(5), 593-602.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res, 36*(6), 1276-1285.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *J Acoust Soc Am, 119*(4), 2455-2466.

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J Acoust Soc Am, 44*(6), 1513-1523.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nat Rev Neurosci, 5*(11), 887-892.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am, 111*(3), 1340-1348.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *Br J Audiol, 35*(3), 173-182.

Gustafsson, H. A., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *J Acoust Soc Am, 95*(1), 518-529.

Harris, K. C., Eckert, M. A., Ahlstrom, J. B., & Dubno, J. R. (2010). Age-related differences in gap detection: effects of task difficulty and cognitive ability. *Hear Res, 264*(1-2), 21-29.

Harris, K. C., Wilson, S., Eckert, M. A., & Dubno, J. R. (2012). Human evoked cortical activity to silent gaps in noise: effects of age, attention, and cortical processing speed. *Ear Hear, 33*(3), 330-339.

Hartley, A. A. (1993). Evidence for the selective preservation of spatial selective attention in old age. *Psychol Aging, 8*(3), 371-379.

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci, 20*(10), 1737-1752.

Helfer, K. S., Chevalier, J., Freyman, R. L. (2010). Aging, spatial cues, and single- versus dual-task performance in competing speech perception. *J Acoust Soc Am, 128*(6), 3625-3633.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear Hear, 29*(1), 87-98.

Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *J Acoust Soc Am, 125*(1), 447-456.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences*. 5th ed. Boston: Houghton Mifflin.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *J Manage, 23*, 723-744.

Hopkins, K., & Moore, B. C. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J Acoust Soc Am, 125*(1), 442-446.

Howes, D. H. (1954). On the interpretation of word frequency as a variable affecting speech of recognition. *J Exp Psychol, 48*, 106-112.

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front Syst Neurosci, 7*(55).

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2016). Corrigendum: Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front Syst Neurosci, 10*(91).

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *J Acoust Soc Am, 120*(5 Pt 1), 2926-2937.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C. & Lee, L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *J Speech Hear Res, 37*(2), 465-474.

Hutka, S. A., Alain, C., Binns, M. A., & Bidelman, G. M. (2013). Age-related differences in the sequential organization of speech sounds. *J Acoust Soc Am, 133*(6), 4177-4187.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objects. *J Acoust Soc Am, 124*(4), 2224-2235.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 19*(6), 741-758.

Jin, S. H., & Nelson, P. B. (2006). Speech perception in gated noise: the effects of temporal resolution. *J Acoust Soc Am, 119*(5 Pt 1), 3097-3108.

Jin, S. H., & Nelson, P. B. (2010). Interrupted speech perception: the effects of hearing sensitivity and frequency resolution. *J Acoust Soc Am, 128*(2), 881-889.

Jin, S. H., Nie, Y., & Nelson, P. (2013). Masking release and modulation interference in cochlear implant and simulation listeners. *Am J Audiol, 22*(1), 135-146.

Johnsurde, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychol Sci, 24*(10), 1995-2004.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J Neurosci, 30*(2), 620-628.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *J Acoust Soc Am, 118*(6), 3804-3815.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *J Acoust Soc Am, 95*(6), 3475-3480.

Kidd, G., Jr., Mason, C. R., & Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J Acoust Soc Am, 114*(5), 2835-2845.

Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *J Acoust Soc Am, 131*(2), 1434-1448.

Krull, V., Humes, L. E., & Kidd, G. R. (2013). Reconstructing wholes from parts: effects of modality, age, and hearing loss on word recognition. *Ear Hear, 34*(2), e14-23.

Kwon, B. J. (2012). Token [Computer program]. Version 1.36, retrieved 5 September 2012 from auditorypro.com/aux/

Kwon, B. J., & Turner, C. W. (2001). Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference? *J Acoust Soc Am, 110*(2), 1130-1140.

Lee, J. H., & Humes, L. E. (2012). Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *J Acoust Soc Am, 132*(3), 1700-1717.

Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *J Exp Psychol Hum Percept Perform, 30*(6), 1077-1091.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci U S A, 103*(49), 18866-18869.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the Neighborhood Activation Model. *Ear Hear, 19*(1), 1-36.

Lunner T. (2003). Cognitive function in relation to hearing aid use. *Int J Audiol, 44*(Suppl 1), 131-143.

Mackersie, C. L., Dewey, J., & Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *J Acoust Soc Am, 130*(2), 1006-1019.

Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *J Assoc Res Otolaryngol, 13*(1), 119-129.

Madix, S. G., Thelin, J. W., Plyler, P. N., Hedrick, M., & Malone J. (2005). *The effects of age and context on phonemic restoration in young adult females.* Poster presented at the American Academy of Audiology Annual Conference, Washington DC.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*(7397), 233-236.

Miller, G. A., Licklder, J. C. R. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am, 22*(2), 167-173.

Molis, M. R., Kampel, S. D., McMillan, G. P., Gallun, F. J., Dann, S. M., Konrad-Martin, D. (2015). Effects of hearing and aging on sentence-level time-gated word recognition. *J Speech Lang Hear Res, 52*(2), 481-496.

Moore, B. C. (1990). Co-modulation masking release: spectro-temporal pattern analysis in hearing. *Br J Audiol, 24*(2), 131-137.

Moore, B. C. (2003). Temporal integration and context effects in hearing. *J Phon, 31*(3-4), 563-574.

Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philos Trans R Soc Lond B Biol Sci, 367*(1591), 919-931.

Moore, B. C., Hafter, E. R., & Glasberg, B. R. (1996). The probe-signal method and auditory-filter shape: results from normal- and hearing-impaired subjects. *J Acoust Soc Am, 99*(1), 542-552.

Moulines, E., & Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commun, 9*, 453-467.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys, 41*(5), 409-415.

Nelson, P. B., & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *J Acoust Soc Am, 115*(5 Pt 1), 2286-2294.

Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J Acoust Soc Am, 113*(2), 961-968.

Newman, R. S., Evers, S. (2007). The effect of talker familiarity on stream segregation. *J Phon, 35*, 85-103.

Oh, S. H., Donaldson, G. S., Kong, Y. Y. (2016). The role of continuous low-frequency harmonicity cues for interrupted speech perception in bimodal hearing. *J Acoust Soc Am, 139*(4), 1747.

Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *J Acoust Soc Am, 125*(1), 457-468.

Peterson, G. H., & Barney, H. L. (1952). Control methods used in a study of the vowels. *J Acoust Soc Am, 24*, 175-184.

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

Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. *Hear Res, 233*(1-2), 114-121.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97*(1), 1197-1205.

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *J Acoust Soc Am, 61*(1), 195-199.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am, 114*(1), 446-454.

R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rajan, R., & Cainer, K. E. (2008). Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience, 154*(2), 784-795.

Rimmele, J., Schröger, E., & Bendixen, A. (2012). Age-related changes in the use of regular patterns for auditory scene analysis. *Hear Res, 289*(1-2), 98-107.

Rönnberg, J. (1990). Cognitive and communicative function: the effects of chronological age and “handicap age”. *Eur J Cogn Psychol, 2*(3), 253-273.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 367-373.

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand J Psychol, 50*(5), 405-418.

Saija, J. D., Akyurek, E. G., Andringa, T. C., & Başkent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *J Assoc Res Otolaryngol, 15*(1), 139-148.

Salthouse, T. A. (2010). *Major Issues in Cognitive Aging.* New York, NY: Oxford University Press.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biol Psychol, 54*(1-3), 35-54.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. J. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology, 14*(1), 102-111.

Samuel, A. G. (1981a). The role of bottom-up confirmation in the phonemic restoration illusion. *J Exp Psychol: Hum Percept Perform, 7*(5), 1124-1131.

Samuel, A. G. (1981b). Phonemic Restoration: insights from a new methodology. *J Exp Psychol: Gen, 110*(4), 474-494.

Scharf, B. (1961). Complex sounds and critical bands. *Psychol Bull, 58*, 205-217.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Percept Psychophys, 42*(3), 215-223.

Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *J Acoust Soc Am, 90*(3), 1332-1339.

Schvartz, K. C., & Chatterjee, M. (2012). Gender identification in younger and older adults: use of spectral and temporal cues in noise-vocoded speech. *Ear Hear, 33*(3), 411-420.

Shafiro, V., Sheft, S., & Risley, R. (2016). The intelligibility of interrupted and temporally altered speech: effects of context, age, and hearing loss. *J Acoust Soc Am, 139*(1), 455-465.

Shafiro, V., Sheft, S., Risley, R., & Gygi, B. (2015). Effects of age and hearing loss on the intelligibility of interrupted speech. *J Acoust Soc Am, 137*(2), 745-756.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Shen, J., Wright, R., & Souza, P. E. (2016). On older listeners’ ability to perceive dynamic pitch. *J Speech Lang Hear Res, 59*(3), 572-582.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn Sci, 12*(5), 182-186.

Shinn-Cunningham, B. G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *J Acoust Soc Am, 123*(1), 295-301.

Simpson, A. N., Matthews, L. J., & Dubno, J. R. (2013). Lipid and C-reactive protein levels as risk factors for hearing loss in older adults. *Otolaryngol Head Neck Surg, 148*(4), 664-670.

Snyder, J. S., & Alain, C. (2005). Age-related changes in neural activity associated with concurrent vowel segregation. *Brain Res Cogn Brain Res, 24*(3), 492-499.

Snyder, J. S., & Alain, C. (2007). Sequential auditory scene analysis is preserved in normal aging adults. *Cereb Cortex, 17*(3), 501-512.

Souza, P., & Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *Int J Audiol, 54*(10), 705-713.

Souza, P., Arehart, K., Miller, C. W., & Muralimanohar (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear 32*(1), 75-83.

Srinivasan, S., & Wang, D. (2005). A schema-based model for phonemic restoration. *Speech Communication, 45*(1), 63-87.

Steinmetzger, K., & Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *J Acoust Soc Am, 138*(6), 3586-3599.

Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *J Acoust Soc Am, 122*(2), 1069-1078.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am, 116*(2), 1081-1091.

Stoltzfus, E. R., Hasher, L., & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In J. E. Richardson, R. W. Engle, L. Hasher, R. H. Logic, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp.66-88). London: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol, 18*(6), 643-662.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res, 28*(3), 455-462.

Summers, V., & Leek, M. R. (1998). FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J Speech Lang Hear Res, 41*(6), 1294-1306.

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *J Neurosci, 32*(5), 1747-1756.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Roverud, E., Kidd, G. Jr. (2016). Role of binaural temporal fine structure and envelope cues in cocktail-party listening. *J Neurosci, 36*(31), 8250-8257.

Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J Speech Hear Res, 35*(6), 1410-1421.

Tan, M. N., Robertson, D., & Hammond, G. R. (2008). Separate contributions of enhanced and suppressed sensitivity to the auditory attentional filter. *Hear Res, 241*(1-2), 18-25.

Tiffin, J., & Asher, E. J. (1948). The Purdue Pegboard: norms and studies of reliability and validity. *J Appl Psychol, 32*(3), 234-247.

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: A comprehensive review. *J Am Geriatr Soc, 40*(9), 922-935.

Trainor, L. J., & Trehub, S. E. (1989). Aging and auditory temporal sequencing: ordering the elements of repeating tone patterns. *Percept Psychophys, 45*(5), 417-426.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *The Stroop Neuropsychological Screening Test.* Odessa, FL: Psychological Assessment Resources.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol Aging, 17*(3), 453-467.

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. http://www.iphod.com.

van Noorden, L. P. A. S. (1975). Temporal Coherence in the Perception of Tone Sequences. Unpublished doctoral dissertation, Eindhoven University of Technology.

van Rooij, J. C. G. M., Plomp, R. & Orlebeke, J. F. (1989). Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery. *J Acoust Soc Am, 86*, 1294-1309.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am, 82*(4), 1152-1161.

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychol Aging, 18*(2), 332-339.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychol Aging, 18*(3), 443-460.

Verschuure, J., & Brocaar, M. P. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. *Percept Psychophys, 33*(3), 232-240.

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am, 107*(3), 1671-1684.

Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *J Acoust Soc Am, 90*(2 Pt 1), 858-865.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain Lang, 68*(1-2), 306-311.

Vongpaisal, T., & Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *J Speech Lang Hear Res, 50*(5), 1139-1156.

Wang, X., & Humes, L. E. (2010). Factors influencing recognition of interrupted speech. *J Acoust Soc Am, 128*(4), 2100-2111.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392-393.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychol Bull, 96*(2), 371-383.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: perceptual synthesis of absent sounds. *Science, 176*(4039), 1149-1151.

White, L. J., & Carlyon, R. P. (1997). Detection of signals having expected and unexpected temporal structures. *Hear Res, 112*(1-2), 141-146.

Woods, K. J. P., & McDermott, J. (2015). Attentive tracking of sound sources. *Curr Biol, 25*(17), 2238-2246.

Wright, B. A., & Dai, H. (1994). Detection of unexpected tones with short and long durations. *J Acoust Soc Am, 95*(2), 931-938.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol Aging, 15*(1), 88-99.

Zekveld, A. A., George, E. L., Kramer, S. E., Goverts, S. T., & Houtgast, T. (2007). The development of the text reception threshold test: a visual analogue of the speech reception threshold test. *J Speech Lang Hear Res, 50*(3), 576-584.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron, 77*(5), 980-991.

Zhang, T., Dorman, M. F., & Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear Hear, 31*(1), 63-69.