Effects of Aging and Noise on Real-Time Spoken Word Recognition: Evidence From Eye Movements

Boaz M. Ben-David
University of Toronto, Mississauga, Ontario, Canada, and Toronto Rehabilitation Institute, Toronto, Ontario, Canada

Craig G. Chambers
Meredyth Daneman
University of Toronto, Mississauga

M. Kathleen Pichora-Fuller
University of Toronto, Mississauga, and Toronto Rehabilitation Institute

Eyal M. Reingold
Bruce A. Schneider
University of Toronto, Mississauga

Purpose: To use eye tracking to investigate age differences in real-time lexical processing in quiet and in noise in light of the fact that older adults find it more difficult than younger adults to understand conversations in noisy situations.
Method: Twenty-four younger and 24 older adults followed spoken instructions referring to depicted objects, for example, "Look at the candle." Eye movements captured listeners' ability to differentiate the target noun (*candle*) from a similar-sounding phonological competitor (e.g., *candy* or *sandal*). Manipulations included the presence/absence of noise, the type of phonological overlap in target–competitor pairs, and the number of syllables.
Results: Having controlled for age-related differences in word recognition accuracy (by tailoring noise levels), similar online processing profiles were found for younger and older adults when targets were discriminated from competitors that shared onset sounds. Age-related differences were found when target words were differentiated from rhyming competitors and were more extensive in noise.
Conclusions: Real-time spoken word recognition processes appear similar for younger and older adults in most conditions; however, age-related differences may be found in the discrimination of rhyming words (especially in noise), even when there are no age differences in word recognition accuracy. These results highlight the utility of eye movement methodologies for studying speech processing across the life span.

KEY WORDS: spoken word recognition, aging, noise, eye movements

A central question in life span approaches to speech communication concerns the effects of aging on adults' ability to interpret spoken language in different situations. Research on this topic has demonstrated that older listeners tend to experience greater difficulty understanding speech compared with younger listeners, particularly in challenging environments such as those with background noise or in which the target talker's speech must be distinguished from the speech of other talkers (for recent reviews, see Humes & Dubno, 2009, and Schneider, Pichora-Fuller, & Daneman, 2009). Several factors have been shown to contribute to these problems. First, age-related changes in the auditory periphery can reduce and/or degrade the quality of information transmitted to the central nervous system, thereby impeding speech understanding, especially in noisy situations (e.g., Divenyi & Haupt, 1997a, 1997b, 1997c; Divenyi, Stark, & Haupt, 2005; Humes, 1996, 2007; Humes, Lee, & Coughlin, 2006). Age-related changes in the central auditory system may affect speech understanding as well (for a review, see Canlon, Illing, & Walton, 2009). Age-related changes in the cognitive functions supporting language processing may further compound comprehension problems in older adults (for reviews, see Wingfield, 1996, and Wingfield & Tun, 2007). Finally, there are interactions between auditory and cognitive processing, especially when listening is effortful (for reviews, see Pichora-Fuller & Singh, 2006, and Schneider & Pichora-Fuller, 2000). Together, these auditory and cognitive factors conspire to have a more deleterious effect on older adults' speech understanding in the complex and noisy situations typical of everyday life, compared with when listening to clear speech in quiet (for a review, see Schneider, Li, & Daneman, 2007).

In noisy situations, older adults, including those with relatively good audiometric thresholds, typically require a higher signal-to-noise ratio (SNR) than younger adults to achieve similar levels of accuracy on measures of word recognition. Adjusting the SNR somewhat offsets the age-related differences due to changes in auditory processing. Indeed, studies that have used SNR adjustments have typically found no age differences in word recognition in low-context sentences or even on more complex tasks involving discourse comprehension (e.g., Gordon, Daneman, & Schneider, 2009; Murphy, Daneman, & Schneider, 2006; Schneider, Daneman, & Murphy, 2005; Schneider, Daneman, Murphy, & Kwong-See, 2000). However, even when the SNR is adjusted so that younger and older adults perform with the same level of accuracy on typical offline measures of word recognition, the possibility remains that the online processes underlying real-time word recognition could differ depending on the SNR and/or on age. Words are not necessarily recognized in the same way in different SNR conditions, and a given level of accuracy in word recognition is not necessarily achieved in the same way by younger and older listeners (for a review, see Pichora-Fuller, 2007).

In the present study, we explored the online processes underlying the recognition of spoken words in older and younger listeners and how these processes are affected by noise. Our primary goal was to characterize the dynamics of the word recognition process in younger and older adults when the level of noise has been adjusted so that both younger and older adults achieve equivalent levels of accuracy when word recognition is measured.

Spoken Word Recognition and the Effects of Aging

Research in the area of spoken word recognition explores how a word form is differentiated from other candidates in the listener's mental lexicon in real time. The cognitive mechanisms that support this process have two key features that relate to the temporal nature of the speech signal. The first key feature is *incrementality,* whereby the process of mapping speech sounds to lexical candidates begins immediately on the basis of the initial sounds in a spoken word and operates continuously in response to the unfolding signal. Incremental processing often allows a listener to identify a word before it has been heard in its entirety. The second key feature is that words in the mental lexicon dynamically *compete* with one another for activation as speech information unfolds in time. The relative success of a given candidate at a selected time point can be understood as a probability reflecting (a) initial activation levels defined by lexical frequency; (b) the degree of match between the candidate and bottom-up information in the phonetic signal; (c) characteristics of the lexical neighborhood (i.e., the number and frequency of the lexical alternatives sharing sound-level characteristics with a given candidate); and (d) contextual information, such as the semantic or grammatical appropriateness of various lexical alternatives (for reviews, see Cutler, 1995, and Miller & Eimas, 1995).

The two key features of the recognition process just described could both be affected by aging, even when performance, measured in terms of recognition accuracy, is equated across age groups. For example, the incremental use of speech sounds to identify an unfolding word could proceed more slowly in older adults, yielding subtle yet measurable delays in the time course of word recognition. The idea that information uptake in incremental word recognition is not fully synchronized with the temporal unfolding of the signal is supported even in studies of younger adults in optimal listening conditions (e.g., Dahan & Gaskell, 2007). The possibility that older adults may have greater difficulty in "keeping up" with the speech signal in this regard would be consistent with the claim that aging is accompanied by a generalized slowing of perceptual and cognitive processes (e.g., Cerella, 1990; Cerella & Hale, 1994; Salthouse, 1991, 1994). The nature of lexical competition might also differ in older and younger adults, with accompanying delays in recognition. For example, researchers have suggested that inhibitory processes become less efficient over the course of aging (e.g., Hasher & Zacks, 1988; Zacks & Hasher, 1994). When applied to spoken word recognition, such an account might predict that the activation (and suppression) of similar-sounding lexical candidates would be less controlled in older listeners. For example, as the word *candle* is heard, a similar-sounding alternative, such as *candy,* might receive greater activation in older adults compared with younger adults. As a consequence, older adults presumably would require additional time to arrive at the decision threshold required to distinguish a target word from similar-sounding alternatives. These age-related differences in inhibition and generalized speed of processing may be further exaggerated by the presence of noise due to increased ambiguity in the identity of masked speech sounds (e.g., Burke & Shafto, 2008).

In contrast to the scenarios we have outlined, it is possible that lexical processing might be comparatively stable over the adult life span. Studies that have examined how word recognition accuracy is influenced by similar-sounding alternatives (called *neighbors*) have shown quite similar patterns in older and younger adults. For example, the increased difficulty that characterizes younger listeners' identification of "hard" words in speech (e.g., low-frequency words with numerous and/or high-frequency neighbors) compared with "easy" words (e.g., higher frequency words with few and/or comparatively low-frequency neighbors; see Luce & Pisoni, 1998) has also been found in older individuals with or without acquired hearing loss and when speech is presented in quiet or in noise (e.g., Dirks, Takayanagi, & Moshfegh, 2001; Dirks, Takayanagi, Moshfegh, Noffsinger, & Fausti, 2001; Sommers, 1996). However, some studies have also reported age-related differences in certain aspects of word recognition. For example, Sommers (1996) and Sommers and Danielson (1999) have found that older individuals perform worse at recognizing hard words even when the performance of younger and older listeners is equated on easy words—a difference attributed to age-related declines in inhibitory control. Thus, the findings are inconsistent regarding whether there are age-related differences in the various components of lexical processing. However, it is important to note that these effects (as well as others mentioned earlier) are based on listeners' conscious poststimulus response about the identity of a word stimulus. Although offline accuracy measures of this sort have provided many valuable insights, they are not ideally suited for studying the more dynamic components of the recognition process that we have highlighted thus far. In the current study, we examined lexical processing from a more online perspective, focusing on the differentiation of a target word from lexical alternatives as the speech signal is encountered in real time.

Overview of the Experiment

For this research, we adopted a variant of the *visual world* technique, in which eye movements are monitored as participants follow spoken instructions related to real or depicted objects (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, the timing and pattern of eye movements to objects in a display have been shown to reflect the incrementality of spoken word recognition as well as the influence of numerous linguistic and contextual constraints on the recognition process (for an overview, see Tanenhaus, Magnuson, Dahan, & Chambers, 2000). In addition, recent research has shown that the saccadic behavior of younger and older adults is comparable (e.g., Pratt, Dodd, & Welsh, 2006), suggesting that this technique can be readily applied to studies involving these different age groups. Our experimental design was modeled to a certain degree on a study of spoken word recognition in younger adults conducted by Allopenna, Magnuson, and Tanenhaus (1998). In their study, listeners followed instructions such as "Click on the candle" by using a computer mouse to select the target in a grid containing a number of clip art images displayed on a computer monitor. These images included, among other things, the target object (e.g., candle) and a "phonological competitor" object that matched either the initial sounds of the target word (e.g., *candy*) or rhymed with it (e.g., *sandal*). Aggregate analysis of listeners' eye movements at the millisecond level reflected temporary consideration of the phonological competitor object (i.e., that object was *not* ultimately selected) as the sounds of the target word were encountered, with an earlier and greater proportion of fixations directed to competitors that shared the same onset sounds compared with competitors that rhymed with the target. The results were interpreted as evidence for the continuous uptake of speech sounds from the unfolding signal, consistent with frameworks such as TRACE (e.g., McClelland & Elman, 1986). A large body of research has now demonstrated that the extent to which listeners direct transient fixations to a competitor object can capture early and unconscious aspects of incremental spoken language interpretation at various levels of linguistic analysis (e.g., Kaiser & Trueswell, 2008; McMurray, Tanenhaus, & Aslin, 2002; Tanenhaus et al., 1995).

We adapted Allopenna et al.'s (1998) study to investigate how age and noise influence the processing of spoken words. As Allopenna et al. did, we manipulated the type of speech competitor accompanying the target object (i.e., whether it matched the onset sounds of the target or rhymed with them). However, we also manipulated three additional factors. First, we investigated the effects of age on lexical competition by testing both older and younger adults. Second, in view of the evidence that older adults experience disproportionate difficulties understanding language spoken in noisy environments, we varied whether the spoken sentences were presented in quiet or against a background of continuous speech-spectrum noise. It is important to note that we set the level of the noise differently for the two age groups to equate the groups on their overall accuracy of word recognition. This ensured that any observed age-related differences in lexical-level processing would not simply be attributable to differences in the accuracy of the two age groups in recognizing individual words. Third and finally, we varied the number of syllables, with some targets and associated competitors being monosyllabic (e.g., *harp–heart, house–mouse*) and others being disyllabic (e.g., *pencil–penguin, candle–sandal*). The difference in the number of syllables influences the amount of overlapping sounds in target–competitor pairs. Therefore, age-related differences in components of word recognition related to the extent of overlap with competitor words could also be tested.

It is important to note that different hypotheses concerning the source of older adults' speech understanding difficulties (e.g., age-related auditory processing deficits vs. age-related cognitive declines such as loss of inhibitory control or generalized slowing) predict different patterns of age-related changes in the dynamics of real-time lexical processing. Even under conditions in which younger and older adults are equally accurate at recognizing target words, an age-related decline in the efficiency of inhibitory processing would likely entail greater competition from similar-sounding words. As noted earlier, this would be expected to delay the point of recognition because of the longer time required to discriminate a target word from alternative lexical candidates. An age-related slowing in neural processing would produce a similar effect on recognition time because of slowed computation of the auditory input. In contrast, a failure to find age-related differences in lexical processing when the performance of younger and older adults is equated for word recognition accuracy (by tailoring noise levels) would be consistent with sensory degradation accounts of older adults' difficulties in speech understanding in noise (i.e., by showing that the adjustment of the SNR can offset age-related differences).

To summarize, in the current study we used eye tracking to look for age-related differences in the processes leading to word recognition. Our goal was to identify some of the factors that contribute to older adults' difficulties in speech understanding, particularly in the noisy situations that often characterize everyday listening.

Method

Participants

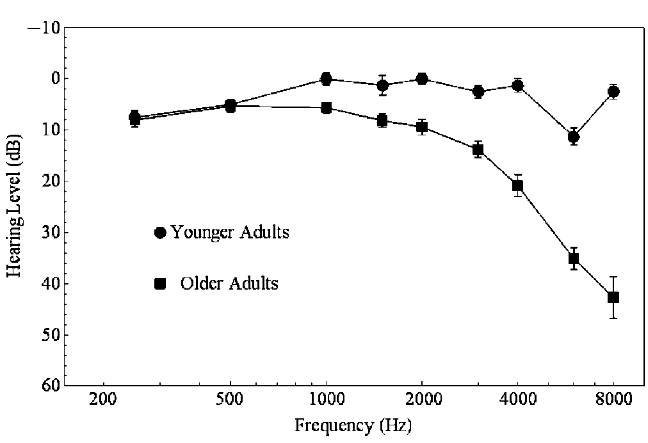
Twenty-four younger adults (M age = 20.2 years, SD = 1.9) and 24 older adults (M age = 70.9, SD = 4.4) participated in the study. The younger adults were undergraduates at the University of Toronto, Mississauga, and received course credit for their participation. The older adults were volunteers from the local community and were paid $10/hr. All participants were native English speakers and achieved a minimum score of 11/20 on the Mill Hill Vocabulary Scale (Raven, 1965), corresponding to normal vocabulary levels for native English speakers. The average Mill Hill score was 13.9/20 (SD = 2.0/20) for younger adults and 15.7/20 (SD = 1.8/20) for older adults. We used a questionnaire to ensure that all participants had good ocular health and no history of visual or auditory pathologies. A Snellen test of visual acuity was conducted for each participant, and participants wore their own corrective eyewear when necessary. All participants had a minimum Snellen fraction within clinically normal limits (20/20 for younger adults, 25/20 for older adults) in the left eye, the right eye (monocular vision), or both (binocular vision), and a minimum score of 7/8 on the Ishihara color blindness test (see Birch, 1997), reflecting normal color vision. The average Snellen fraction for corrected binocular near vision was 13.9/20 (SD = 2.7/20) for younger adults and 21.2/20 (SD = 6.2/20) for older adults. All participants had pure-tone air conduction thresholds within clinically normal limits from 0.25 to 3 kHz in the better ear (≤ 20 dB HL for younger adults and ≤ 25 dB HL for older adults). Hearing levels (averaged over the left and right ears) are shown in Figure 1 for younger and older adults.

Materials

Visual stimuli. There were 32 experimental displays and 32 filler displays. All displays showed four objects. On each trial, object images were presented in the four corners of a 3 × 3 grid displayed on a computer monitor (9 × 9 cm, subtending about 8.5° visual angle horizontally at a distance of 60 cm). Half of the objects depicted monosyllabic nouns, and half depicted disyllabic nouns. Display objects were not recycled on either experimental or filler trials; thus, 256 different images were used in the experiment. The majority of these were drawn from the normed color image set of Rossion and Pourtois (2004). The remaining images were drawn from commercial clip art databases and were selected to match the Rossion and Pourtois images in terms of visual style.

For the experimental displays, we selected 32 pairs of clip art objects. For 16 of the pairs, their respective nouns overlapped in terms of onset sounds (e.g., *tower–towel,*

Figure 1. Mean audiometric pure-tone air conduction thresholds (average of left and right ears) as a function of frequency for the younger and older adults who participated in the study. The vertical bars depict standard errors of the means.

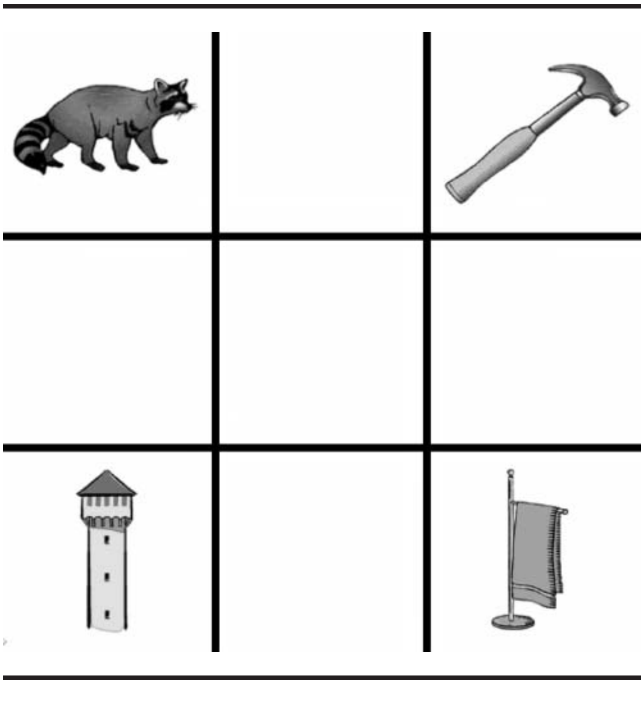


harp–heart). For the remaining pairs, the rhyme portion of the nouns overlapped (e.g., *candle–sandal, house–mouse*). On a given experimental trial, the phonologically related objects were displayed with two unrelated clip art objects whose associated nouns matched the other objects in number of syllables but did not share onset or rhyme sounds (e.g., *hammer* and *raccoon* with *towel* and *tower*; see Appendix A for a complete list of the names of the images used on experimental trials). One of the phonologically related items was identified as the target object (e.g., *towel*) and was referred to in a spoken instruction accompanying the visual display (e.g., "Look at the towel"). The other phonologically related item constituted the competitor object (e.g., *tower*), and it was not mentioned in the instruction. The relative position of object types within the grid (target, competitor, unrelated) was counterbalanced across the set of displays. An example of an experimental trial showing a target–competitor pair with onset overlap is presented in Figure 2.

Filler displays consisted of four clip art objects whose associated nouns did not reflect the type of phonological overlap that characterized the target–competitor pairs used on experimental trials. The fillers were included to prevent participants from developing any expectation that items sharing certain phonological attributes would be mentioned.

Auditory stimuli. To control for potential word frequency differences (as well as potential differences in the recognizability of display objects), the assignment of images to target and competitor objects was reversed across two versions of the experiment (e.g., in one set, *tower* served as the target noun in the spoken instruction, and in the other, *towel* served as the target noun). Each display was presented along with an instruction of the type "Look at the X." Sentences were prerecorded by a female native speaker of Canadian English using a sampling rate of 24414 Hz; the root-mean-square intensity was equated across recordings, and the signal was played at 67.1 dB SPL. To create a second and more challenging listening condition, we subsequently mixed the original recordings with continuous speech spectrum noise. The selected SNR differed between the two groups of participants. For younger adults, we used an SNR of −4 (i.e., noise was presented 4 dB SPL higher than the signal). For older adults, we used an SNR of 0 (the noise and signal were presented at the same level). The use of different SNRs for younger and older adults has been found to result in comparable overall word recognition scores in noise for both age groups (see Murphy, Craik, Li, & Schneider, 2000; Pichora-Fuller, Schneider, & Daneman, 1995). The audio instructions were played from two loudspeakers (Electro-Medical Instrument Co.), each positioned at the height of the listener's head at 45° azimuth.

Figure 2. Black-and-white rendering of a colored display used for an experimental trial. In this example, names corresponding to object images are disyllabic, and the target–competitor pair share acoustic onsets (*towel–tower*).



Procedure

Participants were tested individually in a single-walled sound-attenuating booth. They were seated at a distance of 60 cm from a table-mounted computer monitor, resting their chin on a chin rest. Eye movements were recorded via a table-mounted eye tracking system (Eyelink 2000, SR Research Ltd.), which sampled eye gaze position every 1 ms. Stimuli were presented in blocks according to the presence or absence of background noise. For half of the participants in each age group, the block of noise trials preceded the block of quiet trials, with the order being reversed for the remaining participants.

Each block of trials began with a calibration procedure followed by four practice trials (practice trials for the noise block were also presented in noise). Within each block, 16 experimental trials (eight with onset overlap and eight with rhyme overlap) were randomly interleaved with 16 filler trials, with the exception that the first four trials were always fillers. At the beginning of each trial, a blank grid appeared on the monitor. Participants initiated the presentation of the four clip art objects by pressing a button on a handheld game controller. After 2,000 ms, a short 1-kHz tone was played, directing participants to focus on a black fixation cross that simultaneously appeared in the central square. After the system registered cumulative fixations on the central square for 200 ms, the fixation cross disappeared, and the instruction sentence was played. Participants were instructed to look at, and maintain gaze on, the object denoted in the instruction until a green ("Correct") or red ("Incorrect") square masked that cell (which occurred after 750 ms of cumulative fixations within a cell). The objects then disappeared from the grid to signal the end of the trial.

Note that our procedure differed somewhat from the typical procedure used in prior experiments in which all of the participants were younger adults. We chose to use gaze rather than a mouse click as a response for two reasons. First, there are age-related differences in latency measures when button presses are used as responses (e.g., Cerella & Hale, 1994), whereas age-related differences are minimized when using response measures based on eye movements (e.g., Pratt et al., 2006). Second, in pretests for this study using a "Click on the X" instruction, we found that some older adults shifted their gaze toward the mouse cursor or the actual computer mouse, presumably because of their limited experience in using a computer mouse. Both of these factors would have complicated the analyses had we used the typical mouse click task. In addition, we provided feedback in order to obtain the highest degree of accuracy.

Results and Discussion

Accuracy Data

The mean accuracy data for target selection across experimental conditions and age groups are shown in Table 1. Most generally, these data show that accuracy rates were very high (93.6% overall). Accurate recognition of the target word was at peak when there was no background noise, and recognition was reduced when there was background noise (99.4% vs. 87.8%), particularly when target–competitor pairs had the same onset sounds. Furthermore, noise also appeared to reduce the accuracy of word recognition on trials with monosyllabic target–competitor pairs, but less so on trials with disyllabic pairs. Most important, no evidence of an age-related effect was reflected in the mean accuracy scores in any of the conditions depicted in the table.

Table 1. Mean percentage of trials in which the target object was correctly selected across stimulus conditions and participant groups.

	Quiet		Noise	
	Type of phonological overlap			
Group and word type	Onset	Rhyme	Onset	Rhyme
Younger adults				
Monosyllabic	97.9	100	79.5	91.7
Disyllabic	99.0	99.0	84.4	95.5
M	98.4	99.5	81.9	93.6
Older adults				
Monosyllabic	100	100	79.7	89.6
Disyllabic	99.0	100	87.6	94.4
M	99.5	100	83.7	92.0

We conducted an analysis of variance (ANOVA) with type of competitor (onset or rhyme), acoustic background (quiet or with noise), and number of syllables (monosyllabic or disyllabic) as within-participant factors, and age group (younger or older adults) as a between-participants factor, using word recognition accuracy as the dependent measure. Object set (e.g., whether the target picture was a tower or a towel) and block order (whether the noise block was presented before or after the quiet block) were also included as between-participants factors; however, because these factors were not found to have a significant effect here or in any of the following analyses, they are not discussed further. The analysis revealed a main effect of acoustic background, $F(1, 45) = 104.40$, $p < .001$, $\eta_p^2 = .70$,[1] reflecting higher accuracy for trials in quiet than for trials in noise; a main effect of competitor type, $F(1, 45) = 11.04$, $p < .005$, $\eta_p^2 = .20$, reflecting higher accuracy for rhyme overlap pairs compared with onset overlap pairs; and a marginal main effect of the number of syllables, $F(1, 45) = 3.86$, $p = .055$, $\eta_p^2 = .08$, reflecting higher accuracy for trials containing disyllabic than monosyllabic nouns. We also found a significant interaction of acoustic background and competitor type, $F(1, 45) = 8.6$, $p < .01$, $\eta_p^2 = .16$, reflecting a comparatively larger negative effect of noise when listeners distinguished targets from competitors with overlapping onsets. There was also a significant interaction of acoustic background and number of syllables, $F(1, 45) = 4.38$, $p < .05$, $\eta_p^2 = .09$, reflecting a comparatively larger negative effect of noise in trials with monosyllabic compared with disyllabic target–competitor pairs.

For the current purposes, it is important to stress that accuracy rates did not reliably differ by age group, $F(1, 45) = 0.52$, $p = .8$, and age was not involved in any significant interaction. The absence of an age-related effect on recognition accuracy in quiet is characteristic of performance in ideal listening conditions when no acoustic background noise is presented and listeners have little or no hearing loss (for reviews, see Pichora-Fuller & Souza, 2003, and Schneider & Pichora-Fuller, 2000). Furthermore, the lack of an age-related effect on word recognition in the noise, once the SNR has been adjusted to control for age-related differences in hearing, is consistent with previous work (e.g., Li, Daneman, Qi, & Schneider, 2004) in which age group–specific SNRs were used to compensate for auditory differences between

[1]Partial eta squared (η_p^2) is an estimate of effect size; more specifically, it is the proportion of total variability in the dependent measure that can be attributed to the factor of interest, eliminating other factors from the total nonerror variation, such as the effects of individual differences.

younger and older adults. In summary, the pattern of accuracy scores confirms that our selected SNR levels created a listening task that yielded comparable overall accuracy levels for both age groups. It is important to note that the SNR needed for equivalent speech recognition accuracy in younger and older adults may depend on the materials (see Humes & Dubno, 2009), with older adults needing an SNR between 2 and 6 dB higher than younger adults. The 4-dB difference we used here was based on pilot testing and is similar to the 3 dB typically required when there is little or no supportive sentence context (Pichora-Fuller, 2008).

Calculation of Eye Movement Measures

Although the recognition accuracy measures provide some initial insights into processing differences related to stimulus condition, our primary focus concerned how the pattern and timing of listeners' eye movements reflect the dynamics of the recognition process in cases where the intended target was correctly selected. Consequently, in the calculation of all eye movement measures that follow, we removed from the analysis trials on which participants selected the incorrect display object. We used fixation data from each participant's best eye to calculate fixation position, and we coded fixations that fell within the grid square containing an object as fixations to that object. The timing of eye movements was established relative to the onset of the target noun. The average time interval between the beginning of the recorded sentences and the onset of the target noun (e.g., *towel*) was 898 ms (SD = 85). The average noun duration was 697 ms (SD = 115). The difference in noun duration between monosyllabic target words (690 ms, SD = 114) and disyllabic target words (704 ms, SD = 119), was not significant, $t(62) = 0.5$, $p = .6$. We excluded fixations to display objects beginning earlier than 150 ms after noun onset (0.5% of total trials), to limit the data to fixations that could plausibly be driven by the sounds in the unfolding target noun.

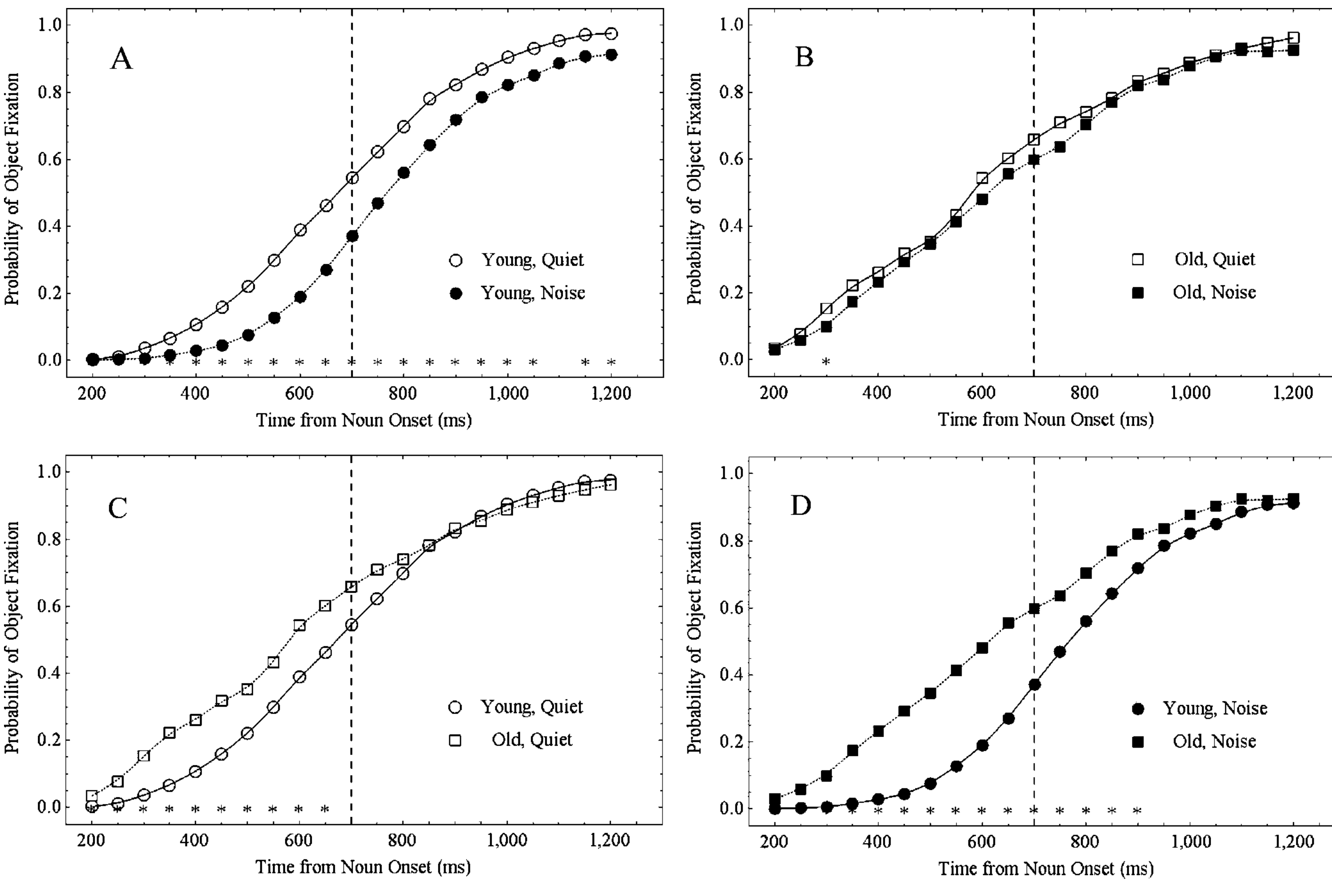
Although the most important eye movement measures are those that reflect listeners' ability to discriminate the target object from the phonological competitor on hearing the target word, we first examined the overall likelihood of fixating *any* display object other than the fixation cross for each age group in quiet and in noise. Because comparatively few visual world studies to date have involved older participants or the presentation of sentences in noise, our goal in conducting this analysis was to provide a general profile of eye movement behavior that could be used to inform analytical or methodological decisions in research using this technique.

The mean likelihood of fixating any display object over time across the quiet and noise conditions is shown separately for the two age groups in the top panels of Figure 3. Although gaze position was sampled once per millisecond, we aggregated the data into 50-ms time bins for ease of analysis and exposition. The analysis region begins 200 ms after the onset of the target word and ends 1 s later (1,200 ms after the onset of the target word). The 200-ms margin built into the analysis window reflects the time typically assumed to reflect the delay between beginning to process speech information and the eye movements that reflect the processing of this information (e.g., see Allopenna et al., 1998). Inspection of the patterns of data suggests that background noise delayed the point at which an eye movement was launched to any of the four display objects in both age groups to a different extent, with a much larger effect on younger adults' fixations. As a statistical test of noise effects, we calculated 95% confidence intervals based on percentiles using a computationally intensive bootstrap resampling method (e.g., Efron & Tibshirani, 1993; van Zandt, 2002). This procedure yields statistical measures that are robust in situations where observed values violate normality and/or reflect different variances across conditions or participant groups. Deviations from normality and unequal variances were often found with the current data set, specifically in comparisons of older and younger adults and when fixation proportions approached 0 or 1.0. Details regarding the application of this procedure to the current data set can be found in Appendix B.

The results of the analysis confirm that the presence of noise caused delays in the time point at which listeners began to launch saccades to display objects and that this effect was not the same for both age groups. For younger adults, the advantage of presentation in quiet began at 300 ms after word onset and persisted until the end of the analysis region (1,200 ms after the onset of the target word). For older adults, the 300-ms mark is the only time interval that shows a significant difference related to the presence of noise. The impact of a noisy acoustic background on the point at which listeners began visual canvassing of the display objects was, therefore, quite small for older adults.

The basis for this difference becomes clear when we replot the data to directly contrast age-related differences within each listening condition (see Figure 3, bottom panels). These data indicate that younger listeners were overall *slower* to begin fixating display objects compared with older listeners. To provide a statistical test of this pattern, we used a between-participants implementation of the bootstrap resampling procedure (see Appendix B). A significant difference across age groups in the 200-ms to 650-ms region (quiet) and the 300-ms to 900-ms region (noise) confirmed the reliability of older adults' overall higher tendency to canvass the displays. We interpreted this result as reflecting a greater degree of uncertainty on the part of older listeners, which manifested itself in terms of increased and more constant saccadic

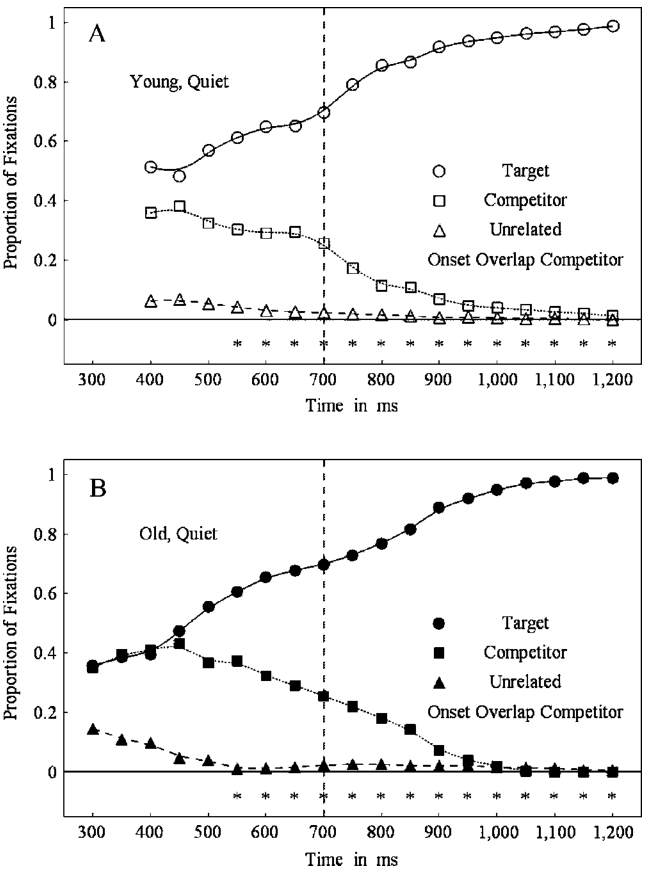
Figure 3. Mean likelihood of fixating any display object over time. The top panels show the effect of acoustic background (quiet vs. noise) for younger adults (Panel A) and older adults (Panel B). In the bottom panels, the data are replotted to show the effect of age in quiet (Panel C) and in noise (Panel D). For this and all subsequent figures, smooth lines were fit by a B-spline function based on a polynomial fit to the data. The vertical line (700 ms) represents average word offset. Asterisks indicate the time intervals at which the difference between the plotted means exceeded 95% confidence intervals.



interrogation of the visual display. The basis for this uncertainty might involve one or more of the following four factors: (a) older adults' more wary response to the experimental scenario (e.g., Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005); (b) a more deliberate or strategic approach to the task of listening to speech; (c) reduced maintenance of representations in working memory (e.g., Bopp & Verhaeghen, 2007); and (d) a reduced perceptual span for effective vision (e.g., Rayner, Castelhano, & Yang, 2009). It is relevant to point out that preliminary work by Sedivy, Demuth, Chunyo, and Freedman (2000) revealed an analogous effect whereby young children showed a higher overall incidence of eye movements to display objects compared with adults (see also Nation, Marshall, & Altmann, 2003).

The implication of the differences in eye movement behavior we just noted is that comparisons across age groups must compensate for differences in the overall tendency to fixate display objects. This can be achieved simply by first recalculating the data in terms of a proportion measure for each participant in each of the conditions, where the numerator represents fixations to a specific object category (either the target, competitor, or unrelated object) and the denominator represents the summed fixations to any object category (e.g., see Dahan & Tanenhaus, 2004). The likelihood of fixating a target by chance (e.g., due to increased scanning) is directly offset in this measure by corresponding increases in the denominator term. As an example, Figure 4 shows the mean proportion of fixations (across participants) to the target noun, the onset overlap competitor, and the average of the two unrelated objects, when speech was presented in quiet. Results for younger and older adults are depicted in the top and bottom panels, respectively, of the figure. Note that the mean at a given time point is plotted only when at least 10% of fixation measurements showed that gaze had been directed away from the central fixation cross. Overall, these data show quite comparable

Figure 4. Mean proportion of fixations to display objects over time from word onset for a sample condition (onset competitors; sentences presented in quiet). Data for younger adults and older adults are shown in the top and bottom panels, respectively.



fixation profiles for younger and older adults. Although this measure still detects older adults' greater tendency to fixate a phonologically unrelated object during the earliest time windows (consistent with their previously noted increased tendency to scan the display), visual consideration of the target and competitor objects, as well as the point at which they are differentiated, is very similar across age groups.

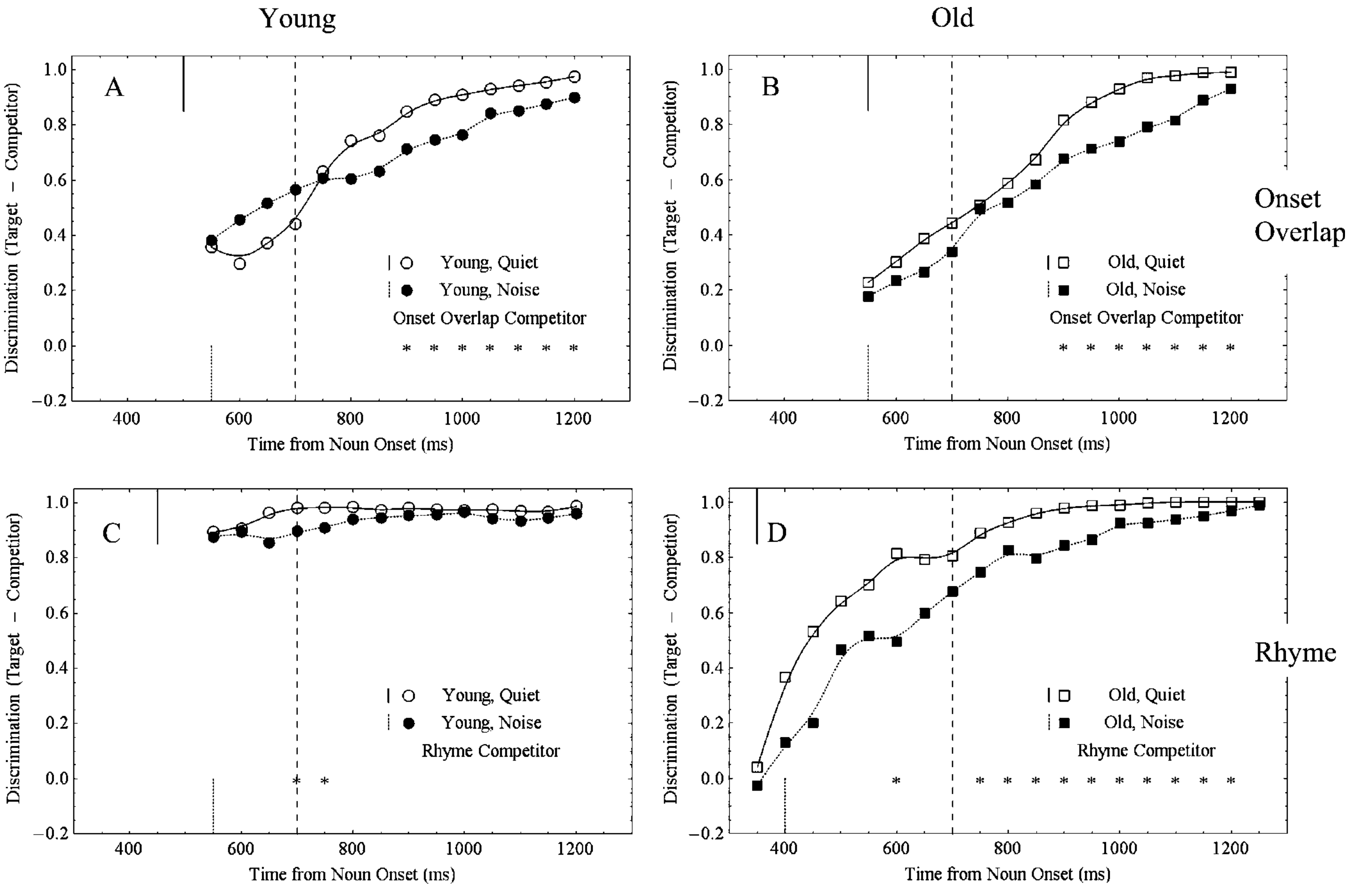
Finally, to combine target and competitor fixation proportions into a single measure that can be used to make direct comparisons across age groups and stimulus conditions, we subtracted the proportion of fixations on the competitor from the proportion of fixations on the target, yielding a *target discrimination score* (for other examples showing the applicability of this dependent measure, see Arnold, Fagnano, & Tanenhaus, 2003; Brown-Schmidt, 2009; and Kaiser & Trueswell, 2008). By this measure, a value near 0 reflects an inability to differentiate the target word from the competitor word at a given time point, whereas a value near 1 reflects full recognition of the target with no consideration of the competitor.

Eye Movement Results

Effects of noise. Differences in listeners' ability to discriminate the target object over time in quiet and in noise are directly contrasted in Figure 5. These differences are plotted separately in the four panels of the figure according to age group (younger vs. older adults) and the type of overlap in target–competitor names (onset vs. rhyme sounds). The short vertical lines perpendicular to the top and bottom margins of each graph indicate the time points at which the 95% confidence interval for target discrimination scores differed from zero, as calculated using the bootstrapping procedure. The condition that corresponds to each line is indicated by whether the line is solid or dashed, as reflected both in the legend and in the plotted discrimination scores. As before, the full dashed horizontal line (around 700 ms) reflects the average offset of the target word, and asterisks denote the time points at which a significant difference was found between the depicted conditions, again using 95% confidence intervals calculated using the bootstrapping procedure (see Appendix B). The mean is plotted only when at least 10% of the observations showed that gaze had been directed away from the central fixation cross in *both* depicted conditions (in this and all following figures).

For younger adults, as indicated by the short lines at the top and bottom margins of Figure 5, the time point at which the discrimination scores significantly differed from zero were earlier in quiet than in noise, both when the competitor shared onset sounds (Panel A) and when rhyme competitors were used (Panel C). However, at later time points, noise had a larger influence on the discrimination of targets from competitors sharing onset sounds. As indicated by the asterisks in Figure 5, noise diminished target–competitor discriminability in onset overlap trials following noun offset for a period of 350 ms (from 900 ms after word onset onward). The comparatively sustained effect of noise in this regard has some similarities with other effects found using the same experimental technique. For example, Dahan and Gaskell (2007) reported that the influence of lexical frequency on words presented in quiet can be found for a considerable period following the words' actual *uniqueness point* (i.e., the point at which a given word can be fully differentiated from alternatives). With rhyme overlap trials, a difference in target discrimination in noise and in quiet was found earlier and lasted only 100 ms (starting at the 700-ms mark). This is due in part to the fact that younger adults' target discrimination scores of rhyming target–competitor pairs in quiet were essentially at ceiling (> 0.9)

Figure 5. Target discrimination scores from word onset, contrasting the effect of presenting sentences in quiet versus in noise. Separate plots show the effect of age (younger vs. older adults, in the left and right panels, respectively) and competitor type (onset vs. rhyme overlap, in the top and bottom panels, respectively). The short vertical lines perpendicular to the top and bottom margins of the graph indicate the time points at which the 95% confidence interval for target discrimination scores differed from zero.



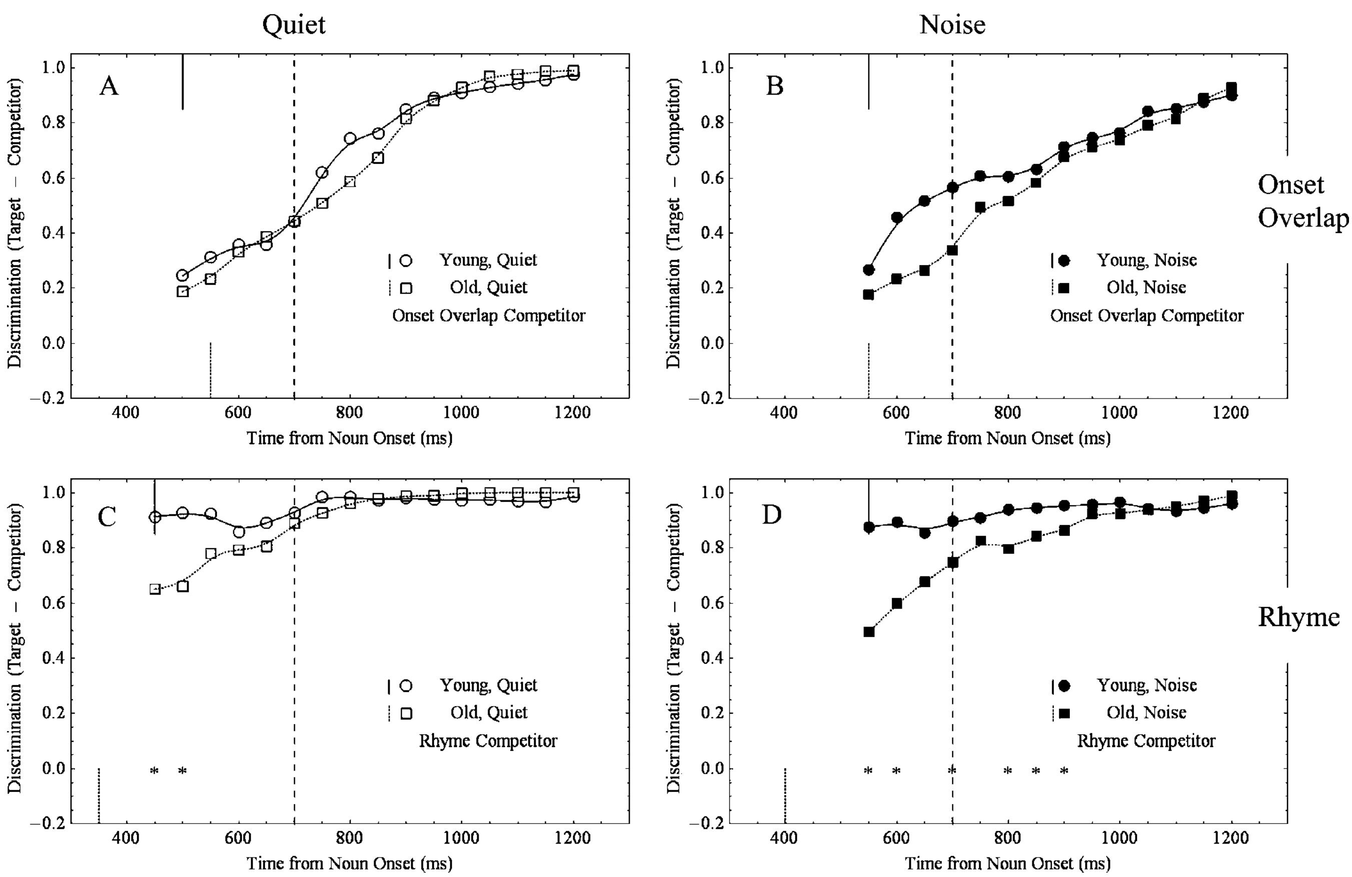
from the 550-ms mark (which reflects processing approximately 350 ms after word onset).[2]

For older adults, the time point at which discrimination scores reliably differed from zero was similar in quiet and noise when targets and competitors shared onset sounds (Panel B of Figure 5); however, noise delayed target discrimination in the rhyme overlap condition (Panel D of Figure 5). As with younger adults, noise continued to have an influence at later time points in trials with onset competitors (from 900 ms after word onset onward); however, for older adults, the effect of noise on target discriminability was also observed at later time points in trials with rhyme competitors, and for a comparatively sustained period (at least 500 ms). An important factor contributing to this difference in the rhyme condition is that the discrimination scores for older adults, unlike those for younger adults, were not at ceiling from the early moments in the unfolding of the target word. Even without background noise, rhyme competitors reflected detectable lexical competition for older adults until the 850-ms mark (i.e., until approximately the end of the word).

Effects of age. The results so far show many instances in which lexical processing behavior was similar

[2]The influence of rhyme competitors in quiet contrasts somewhat with previous studies of younger adults (e.g., Allopenna et al., 1998). This contrast may reflect differences in our implementation of the eye tracking technique and in the method of presentation of the auditory stimuli. First, in the current study, participants indicated their choice of word by using gaze, and they received feedback following each trial. In contrast, in Allopenna et al.'s (1998) study participants clicked on images using a mouse, and no feedback was given. The implementation used here could possibly encourage participants to become more conservative in generating saccades to display objects. Second, the auditory materials presented in the current study were recorded at a sampling rate of 24,414 Hz in studio conditions and were played over high-quality loudspeakers in a single-walled sound-attenuating booth at 67.1 dB SPL. This likely optimized perception of the onset portion of the target word, thereby reducing rhyme competition. In contrast, in Allopenna et al.'s study the auditory materials were read live from a script by an experimenter standing behind the participant in a regular laboratory environment, where the ambient noise level was no doubt higher. It is interesting to note that Allopenna et al.'s results bear a stronger resemblance to the patterns we obtained in the noise condition.

Figure 6. Target discrimination scores from word onset, contrasting the effect of age across acoustic background (quiet vs. noise, in the left and right panels, respectively) and competitor type (onset vs. rhyme overlap, in the top and bottom panels, respectively).



for older and younger listeners, although some differences seem to be found in conditions in which target–competitor pairs shared rhyme sounds. In Figure 6, we have replotted the data shown in Figure 5 to enable direct comparisons of the eye movement profiles of younger and older listeners and permit statistical comparisons of age-related effects on discrimination scores in both onset and rhyme overlap trials.[3] One immediately obvious result is the lack of any significant age-related differences in the conditions using onset competitors (see Panels A and B of Figure 6, for quiet and noise conditions, respectively), as reflected by the lack of asterisks denoting significant contrasts. The first time interval in which the discrimination of target from competitor was significantly greater than zero was also similar regardless of age and noise (500 ms for young:quiet, 550 ms for old:quiet, 550 ms for young:noise, and 550 ms for old:noise).

Figure 6 (see Panels C and D, for quiet and noise conditions, respectively) shows that, for rhyme competitors, the time point at which discrimination scores became significantly different from zero occurred earlier in quiet than in noise for both younger (450 vs. 550 ms) and older (350 vs. 400 ms) adults. In general, younger adults achieved near-asymptotic levels of discrimination shortly after word onset, whereas older adults, primarily in the noise condition, took longer. Because of this, age-related differences in the discrimination of targets from rhyme competitors in quiet emerged 450 ms after word onset and lasted for merely 100 ms. In noise, age-related differences began to emerge a bit later (550 ms after word onset) and persisted for a much longer period of time, up to 900 ms after word onset. Note that it is possible that these age-related differences are underestimated because discrimination scores for younger adults were at ceiling early on.

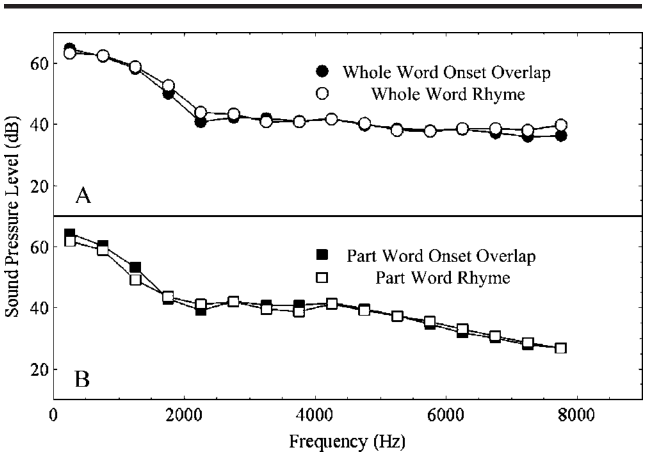
Audibility. Because the target words were presented at the same level to younger and older adults (only the

[3]To conduct within-subject statistical tests for Figure 5, we computed difference scores between the two conditions for each individual. In those cases in which a measure occurred in one but not the other condition, no difference score was computed. In Figure 6, however, the comparison is between groups, and there is no justification for eliminating any of the scores in either of the two conditions when conducting statistical tests. Hence, the data points representing the same condition in Figures 5 and 6 may differ slightly in some instances. A comparison of the two figures suggests that such differences were negligible and affected only the earlier time bins.

level of noise was manipulated), it is possible that the interaction between age and competitor type could reflect spectral differences between the two kinds of target–competitor pairs. To explore this possibility, we plotted the average spectra for rhyme and onset overlap target–competitor pairs in adjacent 500-Hz rectangular bands based on the digital sound files (see Figure 7, Panel A). A repeated-measures ANOVA with frequency bands as a within-participant factor and competitor type (onset overlap vs. rhyme) as a between-participants factor revealed a significant main effect of frequency bands, $F(15, 930) = 227.2$, $p < .001$, $\eta_p^2 = .79$, but no effect of competitor type, $F(1, 62) = 0.09$, $p = .8$, and no interaction between competitor type and frequency bands, $F(15, 930) = 0.59$, $p = .9$. First, note that even though word recognition accuracy was higher in the rhyme condition compared with the onset overlap condition (96.2% vs. 90.9%), the stimulus words in each condition have equivalent spectra. This means that greater accuracy in the rhyme condition cannot be attributed to spectral differences between the stimulus words used in the onset and rhyme overlap conditions. Second, because no spectral differences were found, age-related differences in audibility cannot explain why age-related differences in eye movement measures were found for conditions with rhyme competitors but not for conditions with onset competitors. The logic here is that if age-related differences in audibility are responsible for age-related differences in eye movement patterns for rhymes, they should also produce age-related differences in eye movement patterns in the condition with onset competitors, but there are none.

Because it is possible that the stimulus words in the rhyme and onset competitor conditions might have had different word onset spectra, we also compared the spectra for the first 200 ms of word onset between the two classes of words (see Figure 7, Panel B). Again, in a Frequency Bands × Competitor Type ANOVA, no significant differences were found between rhyme and onset overlap words, $F(1, 62) = 0.11$, $p = .7$. In sum, it is unlikely that age-related differences in audibility can account for the differential impact of age on eye movement measures in onset overlap trials versus rhyme overlap trials.

Figure 7. The average spectra for rhyme and onset overlap target–competitor pairs in adjacent 500-Hz rectangular frequency bands. Panel A shows the average for whole words, and Panel B shows the average for the first 200 ms following word onset.

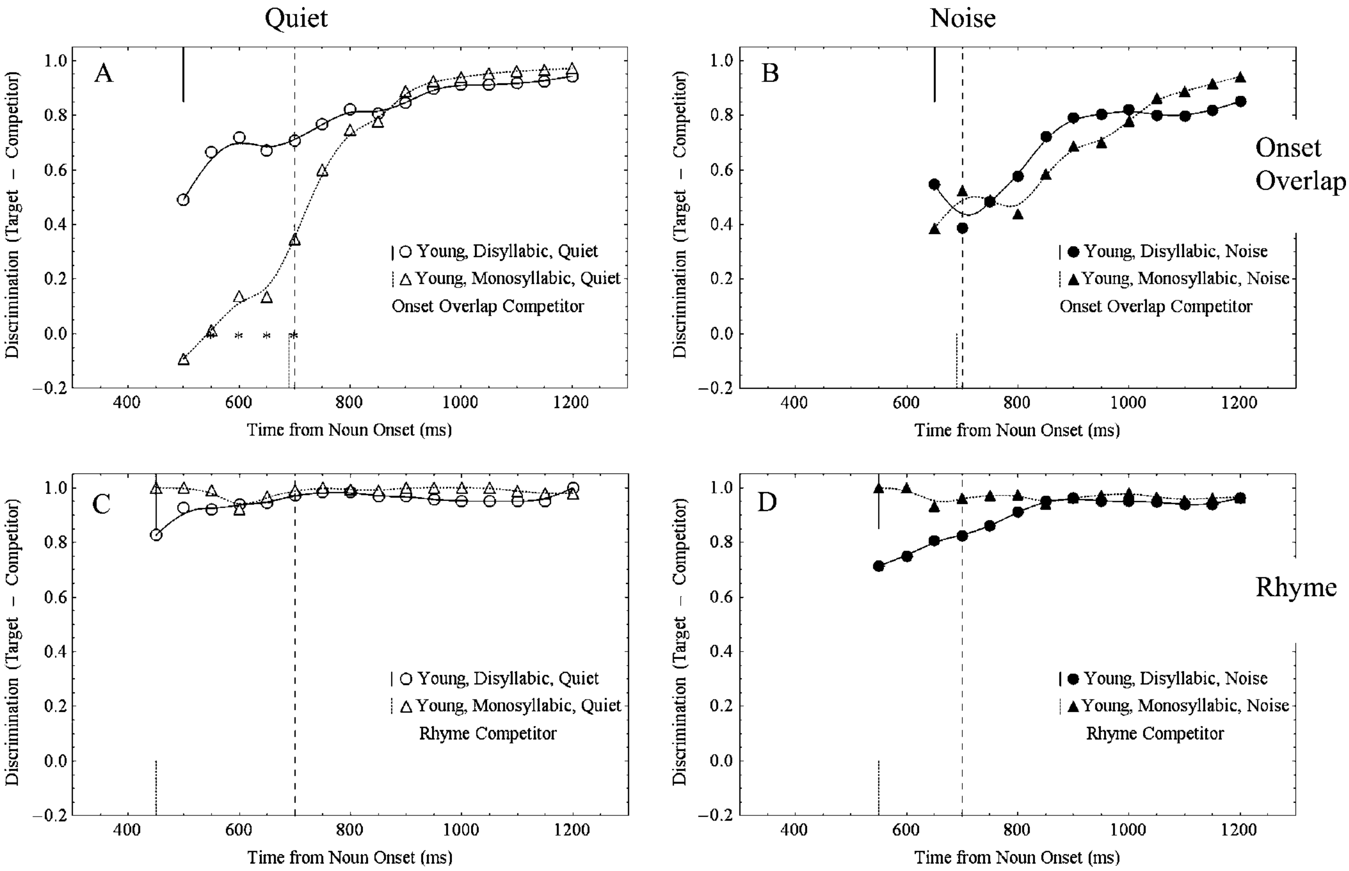


Next, to see whether differences between the two age groups in the audibility of the stimuli had any effect on word recognition accuracy, we tested age-related differences in accuracy for individual word pairs. One could hypothesize that if age-related differences in audibility were a factor, then, even though average error rates were equivalent, there would be some word pairs for which age-related differences in audibility might have led to age-related differences in accuracy. As word pairs contrasted in terms of various speech cues, some contrasts would be preserved and others would be diminished by high-frequency hearing loss in aging (e.g., contrasts based on voiceless stop consonants, such as /p/ vs. /t/). Therefore, we looked for accuracy differences in noise (performance in quiet is at ceiling with 99.5% accuracy) between younger and older adults for each of the onset and rhyme target–competitor pairs using Fisher's exact test[4] (Agresti, 1992). None of these 32 tests were significant ($p > .1$ for all tests). Hence, even at the level of specific target–competitor pairs, we found no evidence that any difference in audibility that might have existed between younger and older adults had any effect on word recognition accuracy.

Effects of the number of syllables. Recall that the names for display objects were monosyllabic for half the experimental trials and disyllabic for the other half. As mentioned earlier, this manipulation had an effect on recognition accuracy scores in the noise condition, whereby monosyllabic words were correctly recognized less often than disyllabic words (for both onset and rhyme competitor conditions and for both younger and older adults). We also examined the eye movement data to explore the influence of the number of syllables. The discrimination of monosyllabic and disyllabic targets by younger listeners is directly contrasted in Figure 8. Separate panels depict the results for the various acoustic background and competitor type conditions. Two noteworthy effects are evident from this figure. First, when the target and the competitor shared *onset* sounds and were played in quiet (see Panel A), discrimination was better with disyllabic compared with monosyllabic targets (as indicated by the asterisks from the 550-ms mark). This is consistent with the fact that the disyllabic target–competitor pairs we used in the onset overlap condition shared a lower

[4]This is a test for statistical significance that is used when sample sizes are small.

Figure 8. Younger adults' target discrimination scores from word onset, contrasting the effect of number of syllables across acoustic background (quiet vs. noise, in the left and right panels, respectively) and competitor type (onset vs. rhyme overlap, in the top and bottom panels, respectively).

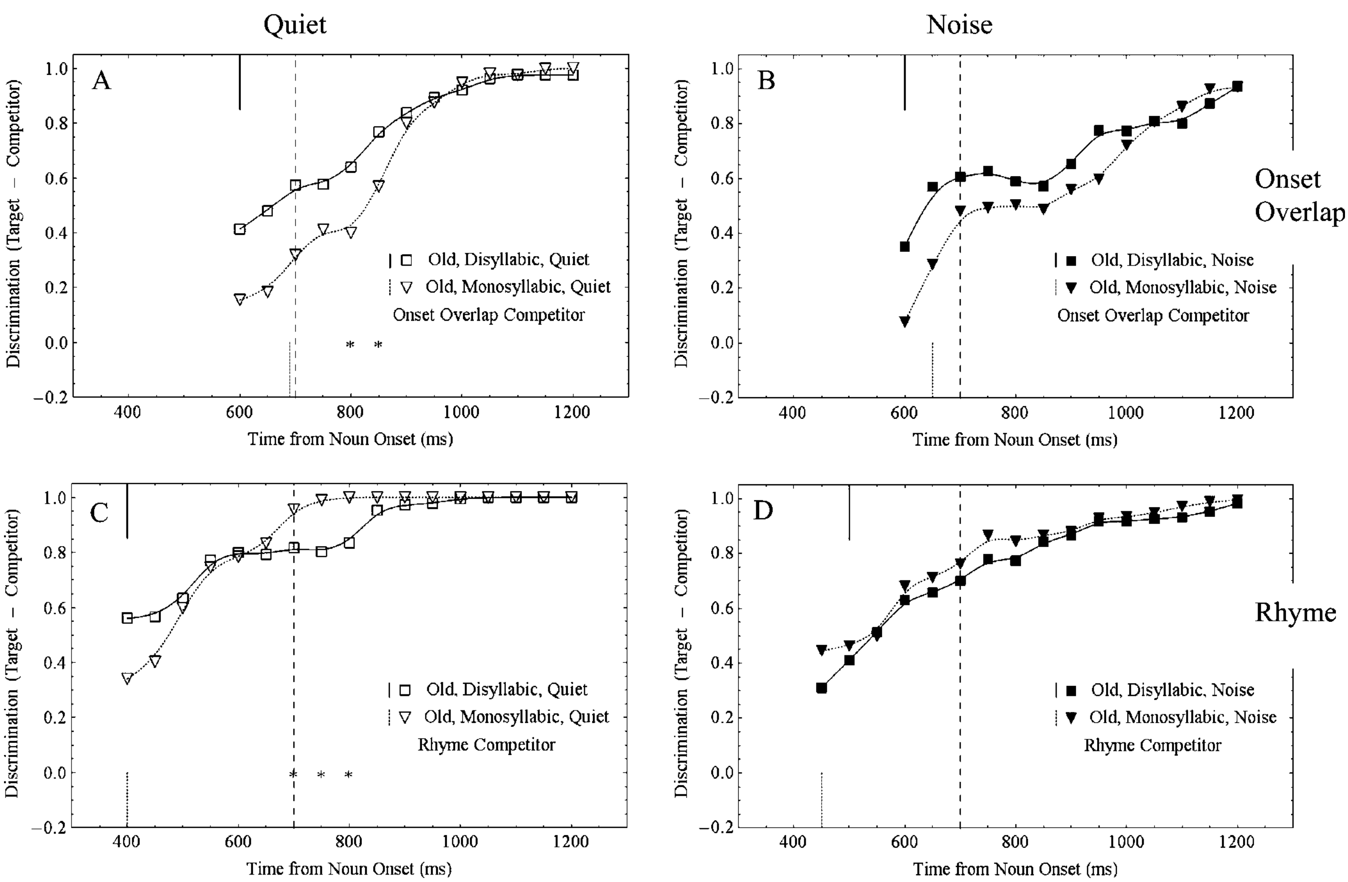


proportion of overlapping phonemes than monosyllabic target–competitor pairs (52% vs. 71% phonemic overlap on average, respectively), leading to an earlier disambiguation point in disyllabic targets. This is reflected in the point at which confidence intervals for the discrimination scores differed from zero (the short vertical lines perpendicular to the top and bottom margins): 500 ms versus 700 ms for disyllabic and monosyllabic targets, respectively. However, this pattern was attenuated when target words with onset competitors occurred in noise (see Panel B).

The second important observation is that the pattern found in the onset overlap condition was almost reversed when target–competitor pairs overlapped in *rhyme* sounds (see Panels C and D of Figure 8); specifically, an advantage for monosyllabic targets was present in the rhyme overlap conditions. Although this outcome was not statistically reliable (note that discrimination scores were at ceiling very early after word onset with disyllabic nouns), it is compatible in principle with the way in which rhyme competitors become activated as the target word unfolds. In particular, even though these competitors (e.g., *nickel*) do not match the target word (e.g., *pickle*) in terms of onset sound(s), evidence supporting their selection accrues as the *remaining* sounds in the target word are heard. When the rhyme portion of the target word consisted of a greater number of speech sounds that overlap with the competitor, as in the disyllabic condition (with 79% phonemic overlap on average), the phonological evidence supporting the competitor was stronger than when the rhyme portion consisted of fewer speech sounds, as in the monosyllabic condition (with 69% overlap).

In Figure 9, we have broken down the data by the number of syllables for the older adult listeners. In general, we see the same patterns as we described earlier for younger adults. With onset overlap trials, older adults' discrimination was better in trials with disyllabic nouns presented in quiet (see Panel A); however, when target–competitor pairs shared rhyme sounds, this pattern was reversed, with better discrimination for monosyllabic words, when presented in quiet (see Panel C). These significant contrasts were not statistically reliable in the noise condition. To summarize, the extent of overlapping sounds in target–competitor pairs (as tested by

Figure 9. Older adults' target discrimination scores from word onset, contrasting the effect of number of syllables across acoustic background (quiet vs. noise, in the left and right panels, respectively) and competitor type (onset vs. rhyme overlap, in the top and bottom panels, respectively).



manipulating the number of syllables) affected eye movement measures similarly across age groups.

Summary and Conclusion

We used a spoken-language eye tracking technique to explore the effects of age, noise, and certain lexical characteristics on the comprehension of spoken words, focusing on dynamic aspects of the recognition process as words unfold in time. We adjusted the SNRs used when presenting the stimuli to compensate for auditory differences in younger and older adults. Listeners' basic accuracy in word recognition was influenced by the type of phonological competitor (onset or rhyme overlap), the number of syllables in the target word (one or two), and the presence of background noise. It is interesting, however, that we found no age-related differences in word recognition accuracy scores in any of the conditions we investigated. Listeners' eye movements, which capture the real-time mechanisms underlying word recognition, reflected the same general pattern of effects found in the accuracy scores. However, the eye movement measures also revealed some intriguing age-dependent differences; specifically, these measures suggest that older adults experience comparatively greater difficulty differentiating target words from rhyming alternatives, primarily when speech is accompanied by background noise. In contrast, real-time processing was similar across age groups when targets had to be distinguished from alternatives sharing onset sounds.

Age-Related Differences in Distinguishing Target Words From Rhyme Competitors in Noise

On trials in which targets occurred with rhyming competitors in a noisy background, eye movements revealed an advantage for younger adults that lasted up to 200 ms after the average word offset. This suggests that, for some words, recognition in noise will proceed more slowly for older adults than for younger adults. When listening to continuous speech, lexical processing in older adults may sometimes be less synchronized to the speech signal than it is in younger adults. Thus, listening to continuous speech possibly requires a greater extent of working memory resources in older than in younger adults, even when they are performing with equal word recognition accuracy. However, the fact that equal word recognition accuracy was maintained, despite slower discrimination in this condition, suggests that older listeners may be able to benefit from their greater experience in language or draw on other compensatory mechanisms.

It is also important to note that the differences between the results obtained in rhyme and onset overlap trials cannot be attributed to possible differences in the audibility of the spoken words because we did not find any differences in the spectra of target words in the two categories, and there were no differences in the spectra of the onset of these words. Furthermore, we did not find any age-related differences in word recognition accuracy, either for the average accuracy rate across stimulus sets or for any single target–competitor pair.

What accounts for the fact that age-related differences were found when distinguishing target words from rhyme competitors in noise, but not when target words were distinguished from onset competitors? One possibility is that older and younger adults may differently weigh informational cues in the process of mapping speech input to alternatives in the mental lexicon. For example, it could be that older adults depend more on amplitude envelope (suprasegmental) cues and less on fine structure (segmental and voice) cues than do younger adults. This explanation would predict that processing a word whose lexical neighborhood contains many rhyming words or words with an otherwise similar amplitude envelope should be more effortful for older adults.

Another possible explanation for the pattern of eye movement data for onset and rhyme competitors hinges on the observation that onset competitors cannot be distinguished at sound onset, whereas rhyme competitors can. If younger adults were faster than older adults at detecting and responding to onset differences, they may differentiate the target word (e.g., *pickle*) from a rhyme competitor (*nickel*) earlier than older adults. Because onset competitors (*towel–tower*) cannot be distinguished at sound onset, younger adults would no longer have this processing advantage. The equivalence in the patterns across age groups for conditions with onset competitors suggests that older adults can compensate for slower initial processing by equally efficient or more efficient processing of the remainder of the word. This perspective is generally consistent with the finding that older adults require greater word onset duration to correctly recognize words than do younger adults (e.g., Craig, 1992; Craig, Kim, Rhyner, & Chirillo, 1993), yet word recognition is facilitated by knowledge of word prosody to the same degree for younger and older adults (Wingfield, Lindfield, & Goodglass, 2000). Furthermore, the fact that older adults achieve the same overall word accuracy when distinguishing targets from rhyme competitors (even in noise) also indicates that, given sufficient time, they can compensate for any slower processing of the initial sound.

Similarities Across Age Groups

Beyond the age-related differences that we detected, our findings suggest that the temporal dynamics of lexical processing are in many ways quite stable across the adult life span. When differentiating target words from alternatives sharing onset sounds, the character of the mechanisms underlying spoken word recognition appears to be quite similar for younger and older adults. In these conditions, older adults' fixations revealed no specific processing delays or detectable increases in lexical competition as compared with younger adults' performance. It is also relevant to point out that the effect of number of syllables on the ability to discriminate the target from the competitor (i.e., the effect of the extent of overlapping sounds) was comparable across age groups in all examined conditions. The proposal that lexical-level mechanisms may be comparatively stable over the course of aging in relation to mechanisms at both "lower" (e.g., sensory- and perceptual-level processing) and "higher" levels (the computation of syntactic and discourse-level relationships) is consistent with a study conducted by Federmeier, Van Petten, Schwartz, and Kutas (2003) that used event-related potentials. Their results showed that although brain responses in older adults showed delays at the sensory and attentional level and at the sentence level (context), the semantic processing of words was not affected. Of course, our study required listeners to recognize single words rather than to process entire sentences or extended discourse. Hence, it is quite possible that there are age-related decrements (e.g., age-related slowing) that are not noticeable in a simple word recognition task but that would emerge in tasks that require the processing of more complex continuous speech.

Taken together, our results are largely consistent with an information-degradation account of age-related declines on cognitive and linguistic tasks (e.g., Lindenberger & Baltes, 1994; Lindenberger & Ghisletta, 2009; Schneider & Pichora-Fuller, 2000; see also Ben-David & Schneider, 2009, 2010, for recent examples regarding visual selective attention), as well as a less efficient uptake of word onset information in older adults. According to this account, apparent declines in performance can arise because the information delivered by the sensory system becomes degraded with age. These declines may be exacerbated by age-related slowing in sensory and cognitive processes.

Finally, we believe that the eye tracking approach we used in this study is likely to provide a highly useful tool for exploring age-related differences in the mechanisms underlying spoken language comprehension. The ability to directly control the visual and auditory context, while preserving the potential for the fine-grained measurement of real-time processing, will enable future studies to explore the online impacts of other auditory, linguistic, and cognitive factors on older listeners' comprehension. These include lexical-level considerations such as word frequency (see Pirog Revill & Spieler, 2009) and detrimental factors such as hearing loss and complex multitalker situations, as well as benefits stemming from clear speech, hearing aid processing, supportive linguistic context, and the use of conversational strategies. Most important, differences in the moment-to-moment processing of unfolding speech that are captured in these measures are often not revealed when traditional offline measures, such as the accuracy of word recognition, are used (see also Pichora-Fuller, 2007).

Acknowledgments

This work was supported by a Strategic Research Training Grant on Communication and Social Interaction in Healthy Aging and a Group Grant on Sensory and Cognitive Aging, both funded by the Canadian Institutes of Health Research (STP-53875 and MGC-42665) and by operating grants from the Canadian Institutes of Health Research (MOP 15359) and the Natural Sciences and Engineering Research Council of Canada (RGPIN 238362-05, RGPIN 2690-02, RGPIN 138472-05, RGPIN 105451-05, RGPIN 9952-03, and RGPIN 9952-08) to the second through sixth authors. The first author was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659).

References

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7,* 131–153.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research, 32,* 25–36.

Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for aging effects in the color–word Stroop task: An analysis of studies. *Aging, Neuropsychology, and Cognition, 16,* 505–534.

Ben-David, B. M., & Schneider, B. A. (2010). A sensory origin for aging effects in the color–word Stroop task: Simulating age-related changes in color-vision mimics age-related changes in Stroop. *Aging, Neuropsychology, and Cognition, 17,* 730–746.

Birch, J. (1997). Efficiency of the Ishihara test for identifying red–green colour deficiency. *Ophthalmic and Physiological Optics, 17,* 403–408.

Bopp, K. L., & Verhaeghen, P. (2007). Age-related differences in control processes in verbal and visuospatial working memory: Storage, transformation, supervision, and coordination. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 62,* P239–P246.

Brown-Schmidt, S. (2009). The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review, 16,* 893–900.

Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 373–443). New York, NY: Psychology Press.

Canlon, B., Illing, R. B., & Walton, J. (2009). Cell biology and physiology of the aging central auditory pathway. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* (pp. 39–74). New York, NY: Springer.

Cerella, J. (1990). Aging and information processing rate. In J. Birren & K. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.

Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life span. *Acta Psychologica, 86,* 109–197.

Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat affect older adults' memory performance. *Experimental Aging Research, 31,* 235–260.

Craig, C. H. (1992). Effects of aging on time-gated isolated word-recognition performance. *Journal of Speech and Hearing Research, 35,* 234–238.

Craig, C. H., Kim, B. W., Rhyner, P. M. P., & Chirillo, T. K. B. (1993). Effects of word predictability, child development, and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research, 36,* 832–841.

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 97–125). New York, NY: Academic Press.

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language, 57,* 483–501.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping form sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513.

Dirks, D. D., Takayanagi, S., & Moshfegh, A. (2001). Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology, 12,* 233–244.

Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., & Fausti, S. G. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing, 22,* 1–13.

Divenyi, P., & Haupt, K. (1997a). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss: I. Age and lateral asymmetry effects. *Ear and Hearing, 18,* 42–61.

Divenyi, P., & Haupt, K. (1997b). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss: II. Correlation analysis. *Ear and Hearing, 18,* 100–113.

Divenyi, P., & Haupt, K. (1997c). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss: III. Factor representation. *Ear and Hearing, 18,* 189–201.

Divenyi, P., Stark, P., & Haupt, K. (2005). Decline of speech understanding and auditory thresholds in the elderly. *The Journal of the Acoustical Society of America, 118,* 1089–1100.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.

Federmeier, K., Van Petten, C., Schwartz, T., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging, 18,* 858–872.

Gordon, M. S., Daneman, M., & Schneider, B. A. (2009). Comprehension of speeded discourse by younger and older listeners. *Experimental Aging Research, 35,* 277–296.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Humes, L. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology, 7,* 161–167.

Humes, L. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology, 18,* 590–603.

Humes, L. E., & Dubno, J. R. (2009). Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* (pp. 211–258). New York, NY: Springer.

Humes, L., Lee, J., & Coughlin, M. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America, 120,* 2926–2937.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes, 23,* 709–748.

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? *Journal of Experimental Psychology: Human Perception and Performance, 30,* 1077–1091.

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 9,* 339–355.

Lindenberger, U., & Ghisletta, P. (2009). Cognitive and sensory declines in old age: Gauging the evidence for a common cause. *Psychology and Aging, 24,* 1–16.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33–B42.

Miller, J. L., & Eimas, P. D. (1995). Speech perception: From signal to word. *Annual Review of Psychology, 46,* 467–492.

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging, 15,* 323–334.

Murphy, D. R., Daneman, M., & Schneider, B. A. (2006). Why do older adults have difficulty following conversations? *Psychology and Aging, 21,* 49–61.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86,* 314–329.

Pichora-Fuller, M. K. (2007). Audition and cognition: What audiologists need to know about listening. In C. Palmer & R. Seewald (Eds.), *Hearing care for adults* (pp. 71–85). Stäfa, Switzerland: Phonak.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*(Suppl. 2), S144–S154.

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

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiological rehabilitation. *Trends in Amplification, 10,* 29–59.

Pichora-Fuller, M. K., & Souza, P. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*(Suppl. 2), S11–S16.

Pirog Revill, K., & Spieler, D. H. (2009). The time course of lexical competition in young and older adults. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 261–266). Austin, TX: Cognitive Science Society.

Pratt, J., Dodd, M., & Welsh, T. (2006). Growing older does not always mean moving slower: Examining aging and the saccadic motor system. *Journal of Motor Behavior, 38,* 373–382.

Raven, J. C. (1965). *The Mill Hill Vocabulary Scale*. London, UK: H. K. Lewis.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging, 24,* 755–760.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33,* 217–236.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2,* 179–183.

Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology, 30,* 240–259.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging, 20,* 261–271.

Schneider, B. A., Daneman, M., Murphy, D. R., & Kwong-See, S. K. (2000). Listening to discourse in distracting settings: The effects of aging. *Psychology and Aging, 15,* 110–125.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18,* 559–572.

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Erlbaum.

Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2009). The effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* (pp. 167–210). New York, NY: Springer.

Sedivy, J. C., Demuth, K., Chunyo, G., & Freeman, S. (2000). Incremental referentiality-based language processing in young children: Evidence from eye movement monitoring. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 684–695). Somerville, MA: Cascadilla Press.

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related changes in spoken word recognition. *Psychology and Aging, 11,* 333–341.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14,* 458–472.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995, June 16). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology: Vol. 4. Methodology in experimental psychology* (3rd ed., pp. 461–516). New York, NY: Wiley.

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

Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, and Hearing Research, 43,* 915–925.

Wingfield, A., & Tun, P. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology, 18,* 548–558.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 241–264). San Diego, CA: Academic Press.

Received October 22, 2009

Revision received April 20, 2010

Accepted June 8, 2010

DOI: 10.1044/1092-4388(2010/09-0233)

Contact author: Boaz M. Ben-David, Oral Dynamics Laboratory, Department of Speech-Language Pathology, 160-500 University Avenue, Toronto, Ontario M5G 1V7, Canada. E-mail: boaz.ben.david@utoronto.ca.

Appendix A. List of the names of the images used on critical trials.

Condition and word type	Target–competitor pairs		Unrelated objects	
Onset overlap				
Monosyllabic	Dog	Doll	Shoe	Broom
	Bone	Bowl	Tie	Pig
	Cloud	Clown	Ring	Wheel
	Knife	Knight	Swan	Tree
	Harp	Heart	Cake	Sun
	Cage	Cane	Drum	Pepper
	Snail	Snake	Pen	Comb
	Card	Cart	Owl	Gun
Disyllabic	Pencil	Penguin	Helmet	Mountain
	Ruler	Rooster	Pumpkin	Basket
	Towel	Tower	Hammer	Raccoon
	Cannon	Candy	Zebra	Finger
	Peanut	Peacock	Iron	Football
	Windmill	Window	Dolphin	Sandwich
	Camel	Camera	Lock	Necklace
Rhyme overlap				
Monosyllabic	House	Mouse	Bow	Nail
	Hat	Bat	Train	Car
	Cake	Rake	Bed	Fish
	Wheel	Seal	Pipe	Fork
	Corn	Horn	Glass	Belt
	Hook	Book	Cow	Ear
	Bear	Pear	Crown	Desk
	Fox	Socks	Key	Ant
Disyllabic	Candle	Sandal	Rabbit	Lemon
	Pickle	Nickel	Anchor	Crayon
	Kitten	Mitten	Apple	Ruler
	Flower	Shower	Monkey	Muffin
	Dragon	Wagon	Bottle	Trumpet
	Carrot	Parrot	Ladder	Guitar
	Paddle	Saddle	Airplane	Toaster
	Racket	Jacket	Suitcase	Button

Appendix B. Bootstrap resampling analysis.

We used a bootstrap resampling procedure to provide tests of significant differences across conditions. The procedure varied slightly according to the type of comparison. See Efron and Tibshirani (1993) and van Zandt (2002) for more in-depth descriptions of the bootstrap technique.

Between-Participant Measures

To test for age-related differences in the likelihood of overall object fixation (see Figure 3), we calculated an object fixation measure for each participant every 50 ms. From the set of 24 scores in each age group at each time point, we then randomly selected, with replacement, a sample of 24 scores. The mean of the sample was calculated, and this entire process was repeated 10,000 times. The resulting set of 10,000 sample means was then ranked, and the 0.025 percentile point and the 0.975 percentile point were identified, allowing us to define a 95% confidence interval (two-tailed) for the difference between the two age groups at each time point. When the limits of the confidence interval for the difference scores were above or below zero for a given time point, an asterisk was used to mark statistical significance. We used a similar procedure for tests of age-related effects in the discrimination scores (see Figures 4, 5, 6, 8, and 9), with the following difference. Because the discrimination score was derived from a proportion measure representing the likelihood of fixating the target, competitor, or one of the unrelated objects, it could yield an undefined value at the earliest time intervals. This could occur because participants might still be fixating the center square (i.e., none of the critical objects), leading to a zero value in the denominator of the proportion score. The starting set of scores in some of the early time bins was therefore restricted to only those participants who had begun fixating objects other than the fixation cross, and we conducted the sampling-with-replacement process using this smaller set of scores.

Within-Participant Measures

The calculation of confidence intervals for within-participant contrasts was similar to the procedure just described, with the exception that the initial phase involved calculating participant-specific difference scores at each time point, based on each individual's "paired" scores in the conditions being compared. These scores constituted the set with which the sampling-with-replacement process was conducted. A significant difference between conditions corresponds to cases in which the boundary for the confidence interval associated with the difference score exceeded zero.