



Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults

Drew J. McLaughlin¹ · Maggie E. Zink² · Lauren Gaunt¹ · Brent Spehar² · Kristin J. Van Engen¹ · Mitchell S. Sommers¹ · Jonathan E. Peelle²

Accepted: 27 July 2021 / Published online: 17 August 2021
© The Psychonomic Society, Inc. 2021

Abstract

In most contemporary activation-competition frameworks for spoken word recognition, candidate words compete against phonological “neighbors” with similar acoustic properties (e.g., “cap” vs. “cat”). Thus, recognizing words with more competitors should come at a greater cognitive cost relative to recognizing words with fewer competitors, due to increased demands for selecting the correct item and inhibiting incorrect candidates. Importantly, these processes should operate even in the absence of differences in accuracy. In the present study, we tested this proposal by examining differences in processing costs associated with neighborhood density for highly intelligible items presented in quiet. A second goal was to examine whether the cognitive demands associated with increased neighborhood density were greater for older adults compared with young adults. Using pupillometry as an index of cognitive processing load, we compared the cognitive demands associated with spoken word recognition for words with many or fewer neighbors, presented in quiet, for young ($n = 67$) and older ($n = 69$) adult listeners. Growth curve analysis of the pupil data indicated that older adults showed a greater evoked pupil response for spoken words than did young adults, consistent with increased cognitive load during spoken word recognition. Words from dense neighborhoods were marginally more demanding to process than words from sparse neighborhoods. There was also an interaction between age and neighborhood density, indicating larger effects of density in young adult listeners. These results highlight the importance of assessing both cognitive demands and accuracy when investigating the mechanisms underlying spoken word recognition.

Keywords Spoken word recognition · Lexical competition · Cognitive load

In activation-competition models of spoken word recognition, candidate words are activated based on incoming acoustic information and compete for recognition against similar-sounding words, or “neighbors” (Luce & Pisoni, 1998; Marslen-Wilson & Welsh, 1978). For a given target word, phonological neighborhood density refers to the number of similar-sounding candidate words in the mental lexicon and can be defined operationally as the number of words that can be created from a target word by adding, deleting, or substitut-

ing a single phoneme. A word such as “scarf” has very few neighbors to compete against (a low-density, or sparse, neighborhood), while a word such as “kit” has many neighbors to compete against (a high-density, or dense, neighborhood). When perceiving a word from a dense neighborhood, listeners must select the correct item from among a greater number of candidates than for words from sparse neighborhoods. Thus, recognizing words from dense neighborhoods is hypothesized to place greater demands on processing than recognizing words from sparse neighborhoods. Importantly, these processes should operate even when word recognition accuracy is perfect (e.g., when listening to speech in quiet).

Evidence in support of activation-competition frameworks comes from many sources. When presented with speech in noise, for example, listeners make more errors recognizing words with many competitors (Dirks et al., 2001; Sommers, 1996; Sommers & Danielson, 1999; Taler et al., 2010). Particularly compelling is the finding that teaching listeners

✉ Drew J. McLaughlin
drewjmclaughlin@wustl.edu

¹ Department of Psychological and Brain Sciences, Washington University in St. Louis, One Brookings Dr, St. Louis, MO 63130, USA

² Department of Otolaryngology, Washington University in St. Louis, St. Louis, MO, USA

new words can increase lexical competition (learning “cathedruke” adds a competitor to “cathedral” after consolidation; Davis et al., 2009; Gagnepain et al., 2012).

Several lines of evidence are consistent with nonlinguistic cognitive processes being recruited to a greater extent for words with many competitors compared with those with fewer. For example, individual differences in cognitive measures related to executive function or inhibition (such as the Stroop task) account for differences in the accuracy of identifying words from sparse versus dense neighborhoods (Dey & Sommers, 2015; Sommers & Danielson, 1999). Functional brain imaging studies suggest a role for inferior frontal gyrus in resolving lexical competition, with greater activity seen for words with larger cohort sizes (Zhuang et al., 2011; Zhuang et al., 2014).

Older adults frequently report challenges with speech understanding. A portion of this challenge may be attributed to age-related hearing loss (Peelle & Wingfield, 2016). However, it has been clear for many years that measures of hearing sensitivity alone are not able to completely account for the difficulties older adults face in speech understanding (Humes et al., 2013), especially for words presented in background noise. Even when speech is fully intelligible, older adults take longer to recognize words (Lash et al., 2013; Wingfield et al., 2000) and show decreased memory for what has been heard (Piquado, Cousins, et al., 2010a; Rabbitt, 1968; Ward et al., 2016), implicating verbal working memory (Cousins et al., 2014). These findings have helped motivate conceptions of effort during listening (Pichora-Fuller et al., 2016), and in particular provided evidence that the cognitive resources needed to understand acoustically challenging speech may be particularly relevant for older adults (Peelle, 2018; Wingfield et al., 2005). Consistent with these broader findings, older adults have more difficulty identifying words with many competitors than do young adults when listening to speech in noise (Sommers, 1996; Sommers & Danielson, 1999). Moreover, this age-related decline in speech recognition for words from dense neighborhoods has been observed when performance for words from sparse neighborhoods is equated across age groups (by individually adjusting signal-to-noise ratios), which implicates cognitive rather than sensory factors (Sommers & Danielson, 1999).

Objectively assessing cognitive demands during successful word recognition can be complicated by a number of factors, including ceiling-level performance when assessing accuracy and individual differences in processing speed when assessing latency (although, see Luce & Pisoni, 1998). These difficulties are often magnified when comparing young and older adults. Indeed, much prior work has examined neighborhood density effects for words presented in noise, to bring performance off ceiling (density effects correspondingly being evident in word recognition accuracy). Pupillometry offers a solution to this problem, and allows for examination of the time course of

cognitive demands. The dilation of the pupil in response to task demands has been found to be a reliable correlate of cognitive processing load across multiple domains (Beatty, 1982; see comprehensive review by Zekveld et al., 2018). For speech perception, research using pupillometry has demonstrated that listeners show a larger pupil response (indicating greater cognitive processing load or “listening effort”) when speech is less intelligible (Zekveld & Kramer, 2014; Zekveld et al., 2010). Additionally, speech that is more difficult to process (due to internal degradation or variation) elicits larger pupil responses even when it is accurately recognized (Winn et al., 2015) and presented in quiet (Brown et al., 2020; McLaughlin & Van Engen, 2020).

However, comparing young versus older adults poses a methodological challenge for pupillometry. Compared with young adults, older adults have smaller resting pupil diameters and smaller pupil responses to light (Bitsios et al., 1996; Guillon et al., 2016). This physiological difference is referred to as *senile miosis*, and may also pose a challenge for comparing the *cognitive* pupil response across age groups (see Piquado, Isaacowitz, et al., 2010b). Going into the present study, our primary concern was that we would find larger pupil responses in young than older adults during word recognition due to physiological differences between the two age groups. To prepare for this possibility, we decided to measure both pupil response to light and the pupil response for a non-speech cognitive task, as points of reference.

In the current study, we used pupillometry to index cognitive load during spoken word recognition in quiet. This work makes a novel contribution to the literature by (1) examining the effect of neighborhood density on cognitive demands (i.e., instead of accuracy or task-based reaction time), (2) examining the time course of (potential) increased cognitive demands for phonologically dense neighborhoods (e.g., are increased demands short- or long-lasting?), and, (3) investigating a variety of perceptual and cognitive abilities (e.g., working memory, processing speed, inhibitory control) that may support word recognition during increased lexical competition. Additionally, unlike prior work in pupillometry, target words were presented in quiet and with no visual competitors (e.g., see Kuchinsky et al., 2013), allowing us to test whether neighborhood density has a measurable effect on the cognitive demand of spoken word recognition even under good listening conditions. Finally, to address these questions in young and older adults, we needed to account for age-related changes in cognitive pupil response, which we did in a novel way.

Methods

Experiment, data, and analysis code are available (<https://osf.io/s5ut8/>).

Participants

Young ($n = 70$) and older ($n = 73$) adult participants were recruited from the Washington University in St. Louis Psychology Participant Pool, the Volunteers for Health research participant registry, and the community under a protocol approved by the Washington University in Saint Louis Institutional Review Board. Subjects recruited from the registry and the community were compensated \$10 per hour for their participation and an additional \$5 for travel expenses (typically \$25 total for 2 hours). Participants from the Psychology Participant Pool received course credit. The young adult group contained subjects 18–25 years old ($M = 19.9$, $SD = 1.7$), and the older adult group contained subjects 65–78 years old ($M = 70.9$, $SD = 3.7$). Forty-four subjects reported that they were male, and 99 reported that they were female. All subjects were native speakers of American English who did not learn a second language before the age of seven. Subjects were pre-screened via a phone interview to ensure that they did not have any diagnosed neurological difficulties or hearing loss requiring a hearing aid. Three young and four older adult subjects were excluded due to data loss in the word recognition pupillometry task, leaving data from 67 young and 69 older adults for the analyses.

The best way to estimate sample size for mixed-effects models (such as with GCM) is via a simulation-based approach. Using simulations, estimates of power to detect an effect can be bootstrapped (e.g., see Green & MacLeod, 2016). However, in order to simulate a power analysis, one needs either prior data or artificial data generated based on a prior effect size. For the present study, neither of these options were available. Thus, we aimed to collect a larger sample size than typically seen in studies of listening effort and pupillometry within practical constraints of time and funding.

Materials

Word recognition task Monosyllabic words in the dense and sparse conditions were matched based on intelligibility (from a pilot in which they were presented in noise at -3 dB SNR to young and older adults), phonemic length, and (as closely as possible) lexical frequency (HAL frequency from the English Lexicon Project; Balota et al., 2007) using MATCH software (Van Casteren & Davis, 2007). Controlling for lexical frequency when examining neighborhood density poses difficulties because the two characteristics co-vary (Baus et al., 2008). In the present study we minimized the differences in lexical frequency between the dense and sparse categories (see Fig. 1b), but differences in frequency could not be eliminated entirely. Neighborhood density was calculated by counting the number of words that differed by a single phoneme (due to addition, deletion, or substitution) from the target word. Words in the sparse category had an average of four

phonological neighbors ($Min = 2$, $Max = 6$) and words in the dense category had an average of 26 phonological neighbors ($Min = 21$, $Max = 34$; see Fig. 1b). Forty words were selected per condition for a total of 80 stimuli. Target words were recorded by a male native speaker of American English in a sound-attenuating booth. The dense stimuli were 618 ms long on average ($SD = 126$ ms) and the sparse stimuli were 628 ms long on average ($SD = 78$ ms).

Procedure

In addition to the primary pupillometry task, in which pupil response was measured during word recognition, two additional pupillometry tasks were included for the specific purpose of dissociating divergences in pupil response due to age-related physiological differences from those due to differences in cognitive processing across the two age groups. Details on these tasks, and all other tests, are provided below.

All subjects provided consent and then completed a battery of tasks and cognitive assessments in the following order: the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), the Shipley Vocabulary Test (Kaya et al., 2012), and the Word Auditory Recognition and Recall Measure (WARRM; Smith et al., 2016). Following these, participants performed three pupillometry tasks measuring (a) individual differences in the range of pupil response to light (light range task), (b) pupil response during an oddball tone discrimination task (oddball tone task), and (c) pupil response during word recognition (word recognition task). Following the pupillometry tasks, participants underwent audiological testing (see Supplemental Fig. S1) and completed a reading span test (Oswald et al., 2015), a Stroop test (Golden & Freshwater, 1978), and a processing speed test.

Cognitive measures The MoCA, Shipley, reading span, and WARRM tasks all followed standard procedures (Golden & Freshwater, 1978; Kaya et al., 2012; Nasreddine et al., 2005; Oswald et al., 2015; Smith et al., 2016). The measure of processing speed was developed in-lab. On each trial, subjects were presented with an image of two colored dots, one blue and one red, positioned near a central white dot (the background was gray). Subjects then pressed one of two keys on the keyboard as quickly as possible (labeled “b” and “r” with tape), indicating which of the dots was closer to the white dot. For half the trials the correct answer was blue and for the other half it was red. Trial order was randomized across subjects. Scores on the processing speed task were the subject’s mean reaction time for correct trials only. 2.6% of trials were excluded for older adult subjects and 1.5% of trials were excluded for young adult subjects.

Audiological testing Hearing tests were conducted in a sound-attenuating booth with a GSI Pello audiometer. For

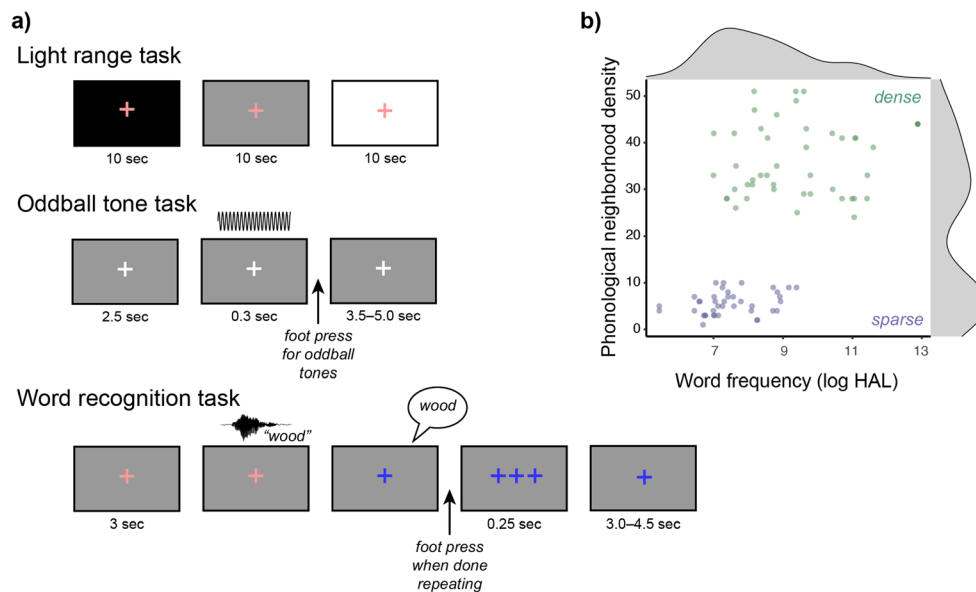


Fig. 1. **a** Schematic of three pupillometry tasks. **b** Word frequency and neighborhood density values for word stimuli

each tested frequency, thresholds were determined by decreasing tone intensity in 10 dB intervals (until the stimulus went undetected by the subject), and then increasing in 5 dB intervals (until the stimulus was detected again); the softest intensity (in 5 dB increments) at which participants could reliably perceive a tone was considered their hearing threshold. Thresholds were recorded at 250, 500, 1000, 2000, 4000, and 8000 Hz for both ears. Participants' better-ear pure tone average (PTA) thresholds were calculated as an average of 500, 1000, and 2000 Hz.¹ Participants also completed the QuickSIN speech-in-noise test (Killion et al., 2004).

Pupillometry A schematic of the pupillometry tasks is shown in Fig. 1a. For all pupillometry tasks, a researcher was present in the room to deliver task instructions, calibrate equipment, and monitor data collection. Between tasks participants were offered short breaks.

Equipment Participants were seated in a sound-attenuating booth with moderate lighting facing a monitor and EyeLink 1000 camera. During the pupillometry tasks, participants rested their chins on a head-mount and wore Sennheiser circumaural headphones. Sound levels were set at a comfortable listening level and did not change between subjects. The EyeLink system recorded pupil area in arbitrary units from the left eye using a 500 Hz sampling rate. Participants wearing glasses were provided lens cleaning wipes. A Chronos foot pedal was used to advance between trials (in place of a button

press) to reduce visual distraction (other researchers anecdotally report participants look at their hands to be sure of correct hand position, causing data loss; J. Reilly, personal communication). Before each task, participants were given verbal and visual instructions and the camera was calibrated to eye movement.

Light range task During the light range task, a fixation cross was presented with alternating black, gray, and white screens, manipulating the luminance at three levels. Each level of luminance was presented for a 10 second interval while pupil area was recorded (repeated three times for a total of nine trials). Participants blinked freely and did not need to make any response during the task.

Word recognition task Each trial of the word recognition task began with a stimulus presentation period and was followed by a verbal response period. The screen showed a red fixation cross during the stimulus presentation period, which included three seconds of quiet preceding and following the target word; the 500 ms immediately preceding the target word was used for baselining. For the response period, the fixation cross changed from red to blue, cueing the participant to repeat the target word aloud. Verbal responses were captured with a recording device for later analysis. After repeating the word aloud, participants stepped on a foot pedal to load the next trial. Here, we inserted a cue in which three blue crosses flashed on the screen (for 250 ms), to confirm that a new trial was loading. Between trials, there was an interstimulus interval of 3,000, 3,500, 4,000, or 4,500 ms (varying randomly) during which the single blue cross remained on screen. The goal of this inter-stimulus interval was to allow time for pupil diameter to return to baseline.

¹ We also calculated and tested a higher range of PTA (1000, 2000, and 4000 Hz), but this composite did not result in different outcomes as compared with the lower range. We elected to use the lower frequency range as this is more common in clinical audiology.

Participants were instructed to reduce their blinking when the fixation cross was red (i.e., during the stimulus presentation period) and blink freely when it was blue. Every eight trials, participants were given mandatory 30 second breaks to prevent fatigue. The task contained 80 trials (40 per condition) presented in random order.

Oddball tone task The oddball tone task was included as a baseline measure to assess age differences in pupil response during discrimination of simple auditory stimuli (and thus, is analyzed independent of the psycholinguistic factors of interest in the current study). The task was intentionally designed to be easier and simpler than the word recognition task. The task included 80 randomly presented trials, 60 of which presented a low-pitch (500 Hz) pure tone and 20 of which presented a high-pitch (2000 Hz) pure tone. The duration of each tone was 300 ms. Subjects were instructed to look at a white fixation cross during the task, and click on the foot pedal as quickly as possible whenever they heard a high-pitched “oddball” tone. Tones were preceded by 2,500 ms of silence on each trial and followed by an interstimulus interval of 3,500, 4,000, 4,500, or 5,000 ms (varying randomly across the experiment). Only the low-pitch trials were retained for analyses; additionally, any low-pitch trials in which a response was made (i.e., false alarms) were discarded. Thus, pupil response on the analyzed (low tone) trials should represent the auditory processing of the tone and inhibition of a response. Given that the tones were easily discriminable, and that the rate of oddball trials was very low, this task should pose very little challenge to adults of all ages.

Data preparation and analysis

Word recognition and oddball tone tasks Pupillometry data was preprocessed in R (Version 3.6.2; R Core Team, 2013; RRID:SCR_001905) using functions from the gazeR package (Geller et al., 2019). Only trials with correct responses were included in analyses. In the word recognition task, older adults accurately reported the target word on 94% of trials and young adults on 96% of trials (95% accuracy across groups). In the oddball tone task, older adults had a 1.5% false alarm rate, and young adults had a 0.3% false alarm rate (a 0.9% false alarm rate across groups). Preprocessing of the data began with identifying, extending, and linearly extrapolating across blinks. Trials missing more than 20% of timepoints were removed, and then subjects missing more than 20% of trials were excluded. After deblinking, the data was smoothed with a 5-point moving average. Baseline pupil size was then calculated and subtracted from all timepoints for every trial within each subject (Reilly et al., 2019). The window for calculating the baseline value was defined as the 500 ms immediately preceding the onset of the target word. Lastly, the pupil data was time binned, reducing the sampling frequency from

500 Hz to 50 Hz. Binning the data lowers computational demand and can reduce temporal autocorrelation in the data (van Rij et al., 2019).

Addressing age-related differences in pupil dynamics It has been demonstrated across multiple studies that older adults have smaller resting pupil diameters (across luminance levels) and smaller pupil responses to light (Bitsios et al., 1996; Guillon et al., 2016), often referred to as senile miosis. Thus, because the same anatomic systems support pupil light response and pupil cognitive response (Mathôt, 2018), prior research comparing young and older adults’ pupil cognitive responses has often normalized these values based on individual subjects’ range in pupil light response (see Piquado, Isaacowitz, et al., 2010b for original use of method).² This transformation of the pupil data increases the pupil size values for subjects with smaller ranges relative to those with larger ranges; that is, it typically increases values for older adults relative to young adults. For the present study, we did not adopt this method. While the formula for normalizing pupil size is used in other studies (e.g., Allard et al., 2010; Ayasse et al., 2016), there is little validation of the formula in the original study or elsewhere. And, some evidence suggests that pupil diameter scales linearly, rather than proportionally, with task demands (Reilly et al., 2019). Instead, we opted to analyze our data without normalization, contextualizing age-related differences in pupil dynamics with an additional measure of pupil cognitive response (oddball tone task; Fig. 2). This decision was motivated by a concern that the restricted range present in pupil light response may not scale directly onto the range present in cognitive pupil response. We mitigated this concern by including a second measure of cognitive pupil response in a nonlinguistic (but auditory) task. However, we also acknowledge that our approach is novel, and return to the topic of how it can be validated and/or improved upon it in the Discussion. Notably, seminal work by Beatty (1982) used a similar tone detection task to examine selective attention. In that task, subjects attended to one pitch (either high or low) and inhibited responses for another. Beatty found greater pupil responses following nontarget tones, indicating a small, but significant, cognitive pupil response when inhibiting a response. Importantly, the size of the pupil response in Beatty’s tone detection experiment was comparable to the size of responses for listening to single words.

In the present study, data from the oddball tone task indicated that the cognitive pupil response in young adults was slightly larger than in older adults (see [Supplemental](#)

² The formula for age-normalized pupil diameter specified by Piquado, Isaacowitz, et al. (2010b) is: $(d_M - d_{\min}) / (d_{\max} - d_{\min})$, where d_M is the pupil diameter (i.e., the pupil cognitive response), and d_{\min} and d_{\max} are the lower and upper bounds of the subject’s pupil light response.

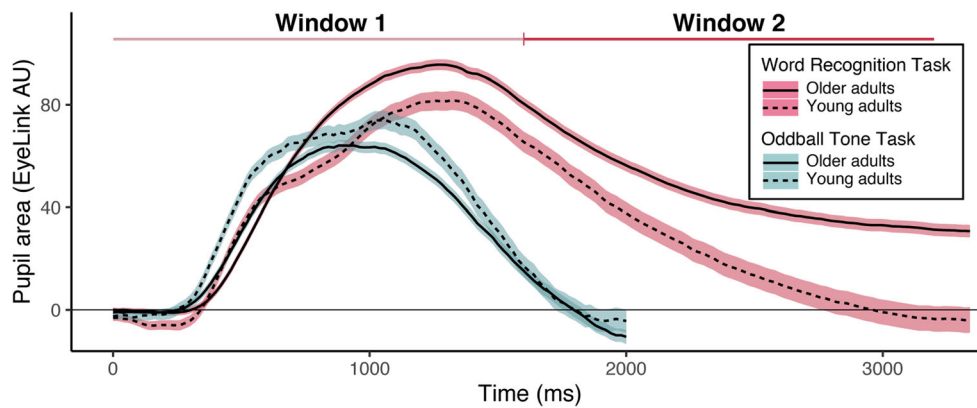


Fig. 2 A comparison of the age effect from the oddball tone task and from the word recognition task. For both tasks, the auditory stimulus started at 0 ms. Boundaries of the analysis windows, which were only used for the word recognition task, are shown above the data

Materials for analyses). This is not particularly surprising; as noted above, prior work has indicated that older adults' pupils may be less responsive (Bitsios et al., 1996; Guillon et al., 2016). Crucially, however, in the context of word recognition we see the opposite outcome (see Fig. 2), with larger pupil response for the older group (discussed further in Results). Together, this data indicates that the size of the effect of age group may be underestimated in the word recognition task (i.e., older adults may have an even larger cognitive demand relative to young adults). However, the benefit of our approach is that it will reduce the likelihood of a Type I error. For the sake of comparison with other studies, we also completed an alternative analysis (see [Supplemental Materials](#)), in which the pupil data was normalized using Piquado et al.'s (2010b) formula. We discuss the different outcomes of these methods in the Discussion.

Model implementation

Here we report the analysis process for the word recognition task (analysis details for the oddball tone task can be found in the [Supplemental Materials](#)). We used growth curve analysis (GCA) to model the pupil response in the word recognition task (Mirman, 2016). The GCA approach uses the lme4 package (Bates et al., 2014) in R to create mixed-effects models that can be shaped to time-course (e.g., pupil response) data. Orthogonalized polynomial terms are implemented as fixed effects in the models, reducing collinearity issues present in traditional polynomial regression.

We identified two analysis windows in the word recognition task: an initial pupil response period (Window 1) and a sustained pupil response period (Window 2; Fig. 2). These windows were defined after viewing a mean pupil response curve (collapsed across all conditions) with a predicted growth curve fit overlaid, to determine if the shape of the model fit was reasonable. The fit to the data is a subjective decision, but was made based on prior experience fitting GCA models. Importantly, it also occurred before any statistical results were

known to the authors, to prevent increasing the risk of Type I error. Window 1 began at trial onset (0 ms) and ended after the peak of the pupil response (1600 ms), resulting in a cubic-shaped response suitable to analysis with polynomial growth curves. We selected data for Window 2 that began immediately after Window 1 (1,601 ms) and continued to 3,200 ms, which captured the majority of the interim period between listening and response, and was also the same length as Window 1 (ensuring equally powered analyses for both windows). Our goal when selecting these analysis windows was to maximize the amount of data included in the analyses while also ensuring that the data could be fit with polynomial growth curves. Although time window selection in pupillometry studies is often at least somewhat arbitrary, a recent analysis suggests that longer time windows increase power to detect effects within a GCA framework (Pelle & Van Engen, 2021). For example, we chose Window 1 to extend slightly past the peak of the pupil response because this allows for a more natural (cubic) fit of the model to the data. We acknowledge that our choice of these analysis windows is partly arbitrary, and that this is a limitation of our analysis.

Here, it should also be noted that our a priori hypotheses and analysis plan included only the initial pupil response period, which we predicted would reflect the cognitive demands of word processing. We decided to explore Window 2 because we suspected that the sustained pupil response may reflect the cognitive demands of maintaining the target words in working memory (i.e., in order to repeat the target word aloud when cued). Thus, we hypothesized that these two areas of interest would separate the cognitive demands of word recognition from those of retention (see Winn, 2016a). However, it would be inaccurate to characterize these two analysis windows as capturing two entirely separate processes. Because of the nature of our experimental design we cannot cleanly disentangle early from late perceptual processing, but our aim was to provide at least some distinction between the initial demands of word recognition and the sustained demands of retaining target items in memory.

Log-likelihood model comparisons were conducted to determine the polynomial shape of each analysis window. Linear, quadratic, and cubic orthogonal polynomials all significantly improved the fit of the Window 1 model and the Window 2 model (all $ps < .05$). We attempted to maximize the random effects structure (Barr et al., 2013), and simplified the structure only as necessary for model convergence. The final random effect structure for both Window 1 and Window 2 included random intercepts by subject and item, and random slopes for the linear and quadratic polynomial terms within subjects (all random effect parameters contributed to model fit, $ps < .05$).

Results

Behavioral

Table 1 shows descriptive statistics of young and older adult performance across all cognitive measures. These values were z -scored before analyses with the pupillometry data. Correlations between the cognitive measures are also shown for young (Table 2) and older adults (Table 3).

Pupillometry

Analysis Window 1 After specifying the polynomial parameters of Window 1, we tested whether including the effects and interactions of age group (factor levels ordered as: young, old) and neighborhood density (factor levels ordered as: sparse, dense) improved model fit using log-likelihood model comparisons. The effect of age group significantly improved model fit, $\chi^2(1) = 5.64$, $p = .02$, and the effect of neighborhood density moderately improved model fit, but was not

significant, $\chi^2(1) = 2.85$, $p = .09$. Model estimates indicated that there was a larger overall pupil response for older compared with young adults ($\beta = 5.70$), and for target words from dense neighborhoods compared with those from sparse neighborhoods ($\beta = 5.60$). The interaction between age group and neighborhood density also improved model fit, $\chi^2(1) = 4.78$, $p = .03$, indicating a larger effect of neighborhood density for young compared with older adults ($\beta = -1.56$). The effects of neighborhood density and age group during Window 1 are visualized in Fig. 3. Next, we examined whether the shape of the pupil response was related to age group and/or phonological neighborhood density by testing models that included interactions of these effects with the linear, quadratic, and/or cubic terms. Age group did not significantly interact with the linear or quadratic polynomial terms (both $ps > .05$), but the interaction with the cubic term did significantly improve model fit, $\chi^2(1) = 4.21$, $p = .04$. For neighborhood density, the interaction with the linear term improved model fit, $\chi^2(1) = 8.94$, $p = .003$, and indicated that words from dense neighborhoods elicited a more rapid increase in pupil size than those from sparse neighborhoods ($\beta = 9.62$). Neither the interaction with the quadratic nor the cubic term improved fit (both $ps > .05$). Lastly, we tested whether individual differences in perceptual and cognitive abilities improved model fit. None of the main effects were significant, with results as follows: better ear PTA, $\chi^2(1) = .73$, $p = .39$, QuickSIN performance, $\chi^2(1) = 1.75$, $p = .19$, processing speed, $\chi^2(1) = .09$, $p = .77$, working memory capacity, $\chi^2(1) = 2.31$, $p = .13$, reading span, $\chi^2(1) = 1.93$, $p = .17$, vocabulary, $\chi^2(1) = .01$, $p = .93$, and Stroop, $\chi^2(1) = 0.25$, $p = .62$. We also tested interactions between each measure and the main effect of age group. Only the measure of processing speed significantly interacted with age group, $\chi^2(1) = 4.95$, $p = .03$. Model estimates indicated that for young adults (the reference level of age group) there

Table 1 Descriptive statistics from the cognitive, linguistic, and perceptual measures of individual subject differences

Measure	Young adults		Older adults		All subjects	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
MoCA	28.46	1.54	26.17	2.44	27.30	2.34
Vocabulary	13.58	1.93	15.04	1.73	14.32	1.97
QuickSIN	0.22	1.14	1.84	2.07	1.04	1.86
Better Ear PTA (dB)	2.64	3.32	13.79	8.09	8.30	8.35
Auditory WMC	4.52	1.13	3.26	1.10	3.88	1.28
Visual WMC	19.16	9.19	12.41	8.72	15.76	9.54
Inhibition (Stroop; ms)	113.23	67.21	219.02	92.51	165.69	96.37
Processing Speed (ms)	627.59	104.88	968.76	243.03	799.44	253.50

Note. Values from measures without units listed represent performance metrics specific to the measure.

MoCA = Montreal Cognitive Assessment; WMC = working memory capacity; PTA = pure tone average threshold; QuickSIN = The Quick Speech-in-Noise Test. Note that lower values indicate better performance on QuickSIN, Better Ear PTA, Inhibition, and Processing Speed.

Table 2 Correlation matrix of the cognitive, linguistic, and perceptual measures of the young adult subjects

Measure	1	2	3	4	5	6	7
1. MoCA	—						
2. Vocabulary	0.34**	—					
3. QuickSIN	-0.17	-0.07	—				
4. Better Ear PTA (dB)	-0.13	-0.29*	-0.2	—			
5. Auditory WMC	0.21	0.23	-0.09	-0.06	—		
6. Visual WMC	0.35**	0.15	-0.13	-0.17	0.48***	—	
7. Inhibition (Stroop; ms)	0.02	-0.26*	-0.1	0.07	0.19	0.13	—
8. Processing Speed (ms)	-0.28*	-0.05	0.19	0.05	-0.29*	-0.09	-0.11

Note. Values represent Pearson correlations (r). Values from measures without units listed represent performance metrics specific to the measure. MoCA = Montreal Cognitive Assessment; WMC = working memory capacity; PTA = pure tone average threshold; QuickSIN = The Quick Speech-in-Noise Test. Note that lower values indicate better performance on QuickSIN, Better Ear PTA, Inhibition, and Processing Speed.

* $p < .05$; ** $p < .01$; *** $p < .001$.

was a positive relationship between processing speed and the overall (mean) size of the pupil response, but for older adults there was no relationship (see Fig. 4). Because higher scores (longer reaction times) on the processing speed task indicate slower processing, a positive relationship with the size of the pupil response indicates that young adult subjects with slower processing also had greater cognitive load during word recognition. We also ran a version of this model after excluding two subjects whose values on the processing speed test were greater than three standard deviations away from the mean, and found the same outcome, though with a slightly smaller effect size, $\chi^2(1) = 3.72, p = .05$. The interactions between age group and better ear PTA, $\chi^2(1) = 2.26, p = .13$, QuickSIN performance, $\chi^2(1) = 0.15, p = .70$, working memory capacity, $\chi^2(1) = .73, p = .39$, vocabulary, $\chi^2(1) = .01, p = .93$, reading span, $\chi^2(1) = .11, p = .75$, and Stroop, $\chi^2(1) = 1.00, p = .32$, did not improve model fit.

Analysis Window 2 After specifying the shape of Window 2, which examined the sustained pupil response caused by retention of the target words in working memory preceding repetition, we tested whether including the effects and interactions of age group (factor levels ordered as: young, old) and neighborhood density (factor levels ordered as: sparse, dense) improved model fit using log-likelihood model comparisons. The effect of age group significantly improved model fit, $\chi^2(1) = 6.49, p = .01$, and the effect of neighborhood density did not improve model fit, $\chi^2(1) = 1.48, p = .22$. Model estimates indicated that there was a larger overall pupil response for older compared with young adults ($\beta = 30.2$). We then examined whether the shape of the pupil response was related to age group and/or neighborhood density by testing models that included interactions of these effects with the linear, quadratic, and/or cubic terms. Age group did not significantly interact with the linear, quadratic, or cubic polynomial terms (all $ps > .05$). Lastly, we tested whether individual differences

Table 3 Correlation matrix of the cognitive, linguistic, and perceptual measures of the older adult subjects

Measure	1	2	3	4	5	6	7
1. MoCA	—						
2. Vocabulary	0.18	—					
3. QuickSIN	-0.28*	-0.23	—				
4. Better Ear PTA (dB)	-0.01	-0.05	0.36**	—			
5. Auditory WMC	0.51***	0.22	-0.51***	0.02	—		
6. Visual WMC	0.25*	0.06	-0.43***	-0.16	0.36**	—	
7. Inhibition (Stroop; ms)	0.07	0.02	0.02	0.08	-0.02	-0.24	—
8. Processing Speed (ms)	-0.14	0.19	0.06	0.05	-0.26*	-0.08	0.18

Note. Values represent Pearson correlations (r). Values from measures without units listed represent performance metrics specific to the measure. MoCA = Montreal Cognitive Assessment; WMC = working memory capacity; PTA = pure tone average threshold; QuickSIN = The Quick Speech-in-Noise Test. Note that lower values indicate better performance on QuickSIN, Better Ear PTA, Inhibition, and Processing Speed.

* $p < .05$; ** $p < .01$; *** $p < .001$.

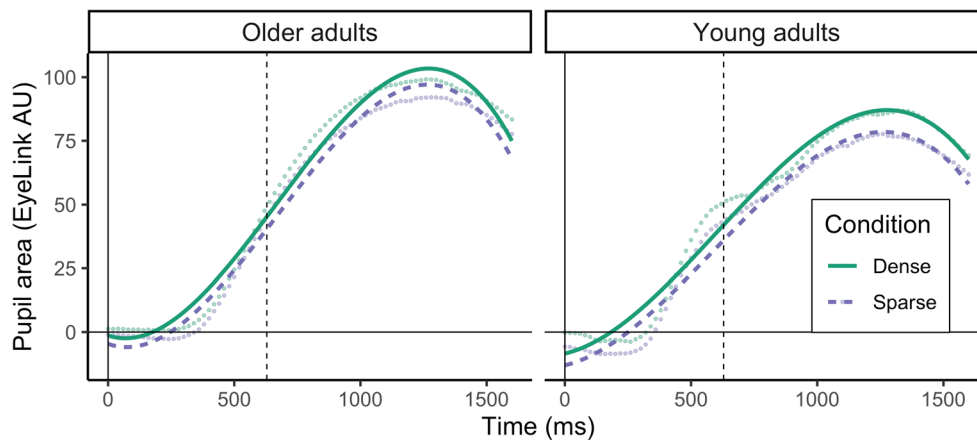


Fig. 3 Model fits of the neighborhood density effect in young and older adults. Mean pupil size is summarized as semitransparent points. For both graphs, the x-axis shows time in milliseconds, with a solid vertical line

marking word onset and a dashed vertical line marking average word offset, and the y-axis shows pupil area in EyeLink arbitrary units (AU)

in perceptual and cognitive abilities improved model fit. None of the main effects were significant, with results as follows: better ear PTA, $\chi^2(1) = 1.03, p = .31$, QuickSIN performance, $\chi^2(1) = 1.57, p = .21$, processing speed, $\chi^2(1) = .31, p = .58$, auditory working memory capacity, $\chi^2(1) = .47, p = .49$, reading span, $\chi^2(1) = .79, p = .37$, vocabulary, $\chi^2(1) = .06, p = .80$, and Stroop, $\chi^2(1) = 0.02, p = .88$. We also tested the interactions between these measures and the effect of age group, and found no significant interactions, with results as follows: better ear PTA, $\chi^2(1) = .15, p = .70$, QuickSIN performance, $\chi^2(1) = .33, p = .57$, processing speed, $\chi^2(1) = 941, p = .33$, working memory capacity, $\chi^2(1) = .86, p = .35$, vocabulary, $\chi^2(1) = .26, p = .61$, reading span, $\chi^2(1) = .32, p = .57$, and Stroop, $\chi^2(1) = 1.05, p = .31$.

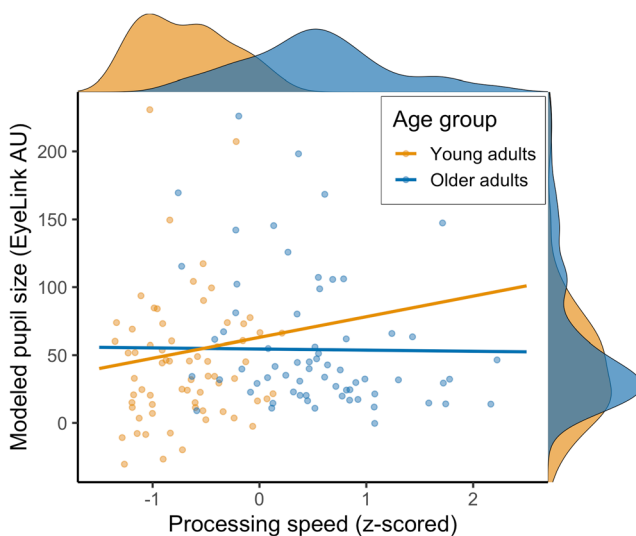


Fig. 4 The interaction between individual differences in processing speed and age group is shown with model fit lines and individual subject points. Density distributions summarize data from young versus older adults for x- and y-axes, respectively. Modeled pupil size (y-axis) is the mean pupil value within the first analysis window

Discussion

In the present study, we used pupillometry to investigate cognitive demands during spoken word recognition for young and older adult listeners. Notably, we specifically examined accurate recognition in quiet. We found that older adults had a larger pupil response during word recognition compared with young adults, indicating greater cognitive demands of speech processing. This result indicates that even word recognition under favorable listening conditions, and in which words are being accurately perceived, imposes a greater cognitive load on older adults as compared with young adults.

We also examined the effect of neighborhood density on cognitive demand during spoken word recognition. Our results indicated a small, marginally significant effect of neighborhood density, with words from dense phonological neighborhoods eliciting a slightly larger overall pupil response than words from sparse phonological neighborhoods. We also found a significant interaction between the effect of neighborhood density and the increase in pupil size immediately following stimulus onset (i.e., during the first window of analysis, Window 1). In other words, this interaction indicates that as the stimulus unfolded, subject pupil response increased at a faster rate for words from dense neighborhoods, compared with words from sparse neighborhoods. However, the increased cognitive demand of processing words with more phonological neighbors does not appear to persist after the presentation of the stimulus; indeed, in an analysis of the pupil response during a time window approximately 1,000 ms after the offset of the stimuli (the “sustained pupil response period”) there was no longer an effect of neighborhood density. This pattern of results is exactly what would be expected, as the initial presentation of the spoken targets is when lexical competition occurs. Altogether, these results indicate that even when word recognition is successful, correctly identifying target words with more lexical competitors may place greater

cognitive demands on the listener, though only during the initial stages of word recognition.

Perhaps surprisingly, our results also indicate that the size of the neighborhood density effect was larger in young adults than in older adults. Prior research has indicated that, when listening to words in background noise, older adults struggle more than young adults with recognizing words from dense neighborhoods (Sommers, 1996; Sommers & Danielson, 1999). Thus, in the present study we expected to see larger density effects in the older adult subjects. One possible explanation for our results is that older adults were already facing greater cognitive demands for word recognition overall (reflected by larger pupil responses generally), and thus were less able to modulate their cognitive effort in response to additional (lexical) challenges.

Currently, there is not a clear consensus on individual factors driving pupil response during speech perception (see Zekveld et al., 2018). For example, some studies have found that poorer hearing predicts larger pupil response (Ayasse et al., 2016; Winn, 2016b), and others have found that poorer hearing predicts smaller pupil response (Kramer et al., 2016; Kuchinsky et al., 2014; Zekveld et al., 2011). Similarly, for cognitive abilities, such as working memory capacity and inhibition, results in both directions have been found (Koch & Janse, 2016; e.g., Koelewijn et al., 2012). Notably, the majority of prior work has examined these individual differences in the context of degraded speech perception, which places additional demands on cognitive systems. In the present study, we examined whether individual differences in hearing ability, working memory, inhibition, processing speed, perception of speech in noise, age-related cognitive impairment, and vocabulary size were related to cognitive pupil response for word recognition in quiet listening conditions. None of these measures of cognitive, linguistic, and perceptual ability significantly predicted variance in pupil response across all subjects. The absence of effects from cognitive measures is consistent with the Ease of Language Understanding (ELU) model of spoken word recognition (Rönnberg et al., 2013; Rönnberg et al., 2008), which proposes that individual differences in cognitive abilities are not predictive of spoken word recognition under favorable listening conditions, such as those used in the current study. It may be that certain abilities only systematically affect listener responses under more challenging conditions (for example, with background noise). However, an interaction indicated that individual differences in processing speed were related to pupil size for young, but not older, adults. Young adult subjects who had poorer (slower) performance on a visual processing speed task also showed larger pupil responses for overall word recognition. This result indicates that young adults

with slower processing speed also recruit greater cognitive resources during spoken word recognition.

The lack of an effect of hearing sensitivity on pupil response was particularly puzzling. In addition to interfering with intelligibility, poorer hearing sensitivity (operationalized as better-ear pure tone averages) has been found to affect sentence comprehension (Wingfield et al., 2006) and memory (Koeritzer et al., 2018; McCoy et al., 2005). One possibility is that the degree of hearing loss in our older adult participants, which was mild-to-moderate sloping hearing loss (see [Supplemental Fig. S1](#)), was not severe enough to sufficiently challenge speech perception in quiet conditions, or that the presentation level of the stimuli (65 dB) was sufficiently high to mitigate any effects of mild to moderate hearing loss in our older adult participants. It would be informative to further investigate lexical competition effects in listeners with poorer hearing.

Notably, in the current study we included a non-linguistic control task to help us interpret group differences in the main task. When designing the study, we were initially concerned that for word recognition we would find larger pupil response in young than older adults (due to senile miosis). Thus, in addition to assessing pupil response during word recognition, we included a cognitive pupillometry task with nonlinguistic auditory stimuli (pure tones) in order to assess cognitive pupil response during a simple auditory task. This measure gave us another point of reference for how anatomical differences may affect the cognitive pupil response. Indeed, in this oddball tone task, we found that young adult subjects had a larger pupil response to the auditory stimuli, which we believe reflects age-related anatomical differences in pupil dynamics. By comparison, in the word recognition task we found larger pupil response in older adults. Ultimately, the age effect in the oddball tone task puts the age effect in the word recognition task into perspective, and indicates that the size of the effect found in the word recognition task—while in the predicted direction—may nonetheless be underestimated due to physiological differences between the two age groups.

We believe that the approach of including a cognitive pupillometry comparison task may prove useful for other researchers examining pupil response across subject groups, particularly those with suspected or known physiological differences in their pupil response. One extension would be to use the comparison task to statistically control for differences between the two groups of interest, either as a covariate or to help estimate a subject-specific pupil response function (Denison et al., 2020). In the present study, we chose to use an oddball tone task because it is auditory but non-linguistic, and Beatty (1982) used a similar tone detection task. However, for work outside of speech perception there may be other tasks that are better suited as comparison measures. Validating and standardizing this novel method for controlling

for age-related physiological differences will be an important next step for pupillometry research.

Importantly, we also completed an alternative analysis of the word recognition task using a scaling method proposed by Piquado et al. (2010b; Supplemental Materials). In this analysis approach, the cognitive pupil data is normalized using subjects' range in pupil response to light. The outcomes of this analysis broadly mirrored the findings of our main analysis (which was conducted on unnormalized pupil data). However, the sizes of the age group effect and the (within-subject) density effect were dramatically greater in the light-range-normalized analysis. Indeed, as discussed above, it is possible that our primary analysis (with unnormalized data) may have underestimated the size of the age effect. However, the converse of this issue is present in the light-range-normalized analysis, in which—crucially—it is possible that the size of the age effect was overinflated (posing the risk of a Type I error).

In summary, the present study demonstrated effects of aging and lexical competition on cognitive demands during spoken word recognition. Using pupillometry as an index of cognitive load, we found that recognizing words from dense phonological neighborhoods may place greater cognitive demands on the listener than recognizing words from sparse phonological neighborhoods. Additionally, even under quiet listening conditions, older adult listeners appear to face greater cognitive load for word recognition than young adult listeners.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-021-01991-0>.

Acknowledgements This work was supported by grant R01 DC014281 from the U.S. National Institutes of Health, and a Graduate Research Fellowship DGE-1745038 from the National Science Foundation.

Open practices statement: The data and materials for all experiments are available (<https://osf.io/s5ut8/>). None of the experiments were preregistered.

References

- Allard, E. S., Wadlinger, H. A., & Isaacowitz, D. M. (2010). Positive gaze preferences in older adults: Assessing the role of cognitive effort with pupil dilation. *Neuropsychology, Development, and Cognition: Section B, Aging, Neuropsychology and Cognition*, *17*(3), 296–311.
- Ayasse, N. D., Lash, A., & Wingfield, A. (2016). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Frontiers in Aging Neuroscience*, *8*, 329.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. In *arXiv [stat.CO]*. arXiv. <http://arxiv.org/abs/1406.5823>.
- Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, *23*(6), 866–888.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.
- Bitsios, P., Prettyman, R., & Szabadi, E. (1996). Changes in autonomic function with age: A study of pupillary kinetics in healthy young and old people. *Age and Ageing*, *25*(6), 432–438.
- Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. (2020). Author accepted manuscript: Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology*, *73*(9), 1431–1433. <https://doi.org/10.1177/1747021820916726>
- Cousins, K. A. Q., Dar, H., Wingfield, A., & Miller, P. (2014). Acoustic masking disrupts time-dependent mechanisms of memory encoding in word-list recall. *Memory and Cognition*, *42*, 622–638.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*, 803–820.
- Denison, R. N., Parker, J. A., & Carrasco, M. (2020). Modeling pupil responses to rapid sequential events. *Behavior research methods*, *52*(5), 1991–2007.
- Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychology and Aging*, *30*(3), 634–646.
- Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., Fausti, & Stephen, A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing*, *22*(1), 1–13.
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology: CB*, *22*, 615–621.
- Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behavior research methods*, *52*(5), 2232–2255.
- Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.
- Golden, C. J., & Freshwater, S. M. (1978). Stroop color and word test.
- Guillon, M., Dumbleton, K., Theodoratos, P., Gobbe, M., Wooley, C. B., & Moody, K. (2016). The effects of age, refractive status, and luminance on pupil size. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *93*(9), 1093–1100.
- Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Frontiers in Systems Neuroscience*, *7*, 55.
- Kaya, F., Delen, E., & Bulut, O. (2012). Test review: Shipley-2 Manual. *Journal of Psychoeducational Assessment*, *30*(6), 593–597.
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4, Pt. 1), 2395–2405.

- Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, 139(4), Article 1618. <https://doi.org/10.1121/1.4944032>
- Koelwijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, 2012, Article 865731. <https://doi.org/10.1155/2012/865731>.
- Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *Journal of Speech, Language, and Hearing Research: JSLHR*, 61(3), 740–751.
- Kramer, S. E., Teunissen, C. E., & Zekveld, A. A. (2016). Cortisol, chromogranin A, and pupillary responses evoked by speech recognition tasks in normally hearing and hard-of-hearing listeners: A pilot study. *Ear and Hearing*, 37(Suppl. 1), 126S – 35S.
- Kuchinsky, S. E., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort: Changes in word recognition and effort. *Psychophysiology*, 51(10), 1046–1057.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr, Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, 50(1), 23–34.
- Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. *Experimental Aging Research*, 39, 235–253.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36.
- Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29–63.
- Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1(1), 16. <https://doi.org/10.5334/joc.18>
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology*, 58(1), 22–33.
- McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupil response for accurately recognized accented speech. *The Journal of the Acoustical Society of America*, 147(2), EL151–EL156.
- Mirman, D. (2016). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC Press.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699.
- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, 47(4), 1343–1355.
- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, 39(2), 204–214.
- Peelle, J. E., & Van Engen, K. J. (2021). Time stand still: Effects of temporal window selection on eye tracking analysis. *Collabra: Psychology*, 7(1), 25961.
- Peelle, J. E., & Wingfield, A. (2016). The neural consequences of age-related hearing loss. *Trends in Neurosciences*, 39(7), 486–497.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S–27S.
- Piquado, T., Cousins, K. A. Q., Wingfield, A., & Miller, P. (2010a). Effects of degraded sensory input on memory for speech: Behavioral data and a test of biologically constrained computational models. *Brain Research Bulletin*, 1365, 48–65.
- Piquado, T., Isaacowitz, D., & Wingfield, A. (2010b). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47(3), 560–569.
- R Core Team. (2013). R: A language and environment for statistical computing. <https://www.r-project.org/>
- Rabbitt, P. M. A. (1968). Channel capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, 20, 241–248.
- Reilly, J., Kelly, A., Kim, S. H., Jett, S., & Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behavior Research Methods*, 51(2), 865–878.
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 31. <https://doi.org/10.3389/fnsys.2013.00031>
- Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *International Journal of Audiology*, 47 Suppl 2, S99–S105.
- Smith, S. L., Pichora-Fuller, M. K., & Alexander, G. (2016). development of the word auditory recognition and recall measure: A working memory test for use in rehabilitative audiology. *Ear and Hearing*, 37(6), e360–e376.
- Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, 11(2), 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(3), 458–472.
- Taler, V., Aaron, G. P., Steinmetz, L. G., & Pisoni, D. B. (2010). Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, 65B, 551–560.
- Van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavioral Research Methods*, 39(4), 973–978.
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, 23. <https://doi.org/10.1177/2331216519832483>
- Ward, C. M., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2016). Effects of age, acoustic challenge, and verbal working memory on recall of narrative speech. *Experimental Aging Research*, 42(1), 126–144.
- 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: JSLHR*, 43(4), 915–925.
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology*, 17(7), 487–497.
- Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*, 14, 144–148.
- Winn, M. (2016a). Pupillary responses signify more than just effort: Windows into processing, prediction, reflection, and uncertainty.

- The Journal of the Acoustical Society of America*, 140(4), 3045–3045.
- Winn, M. B. (2016b). Rapid Release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216516669723>
- Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). the impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, 36(4), e153–e165.
- Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, 22, 2331216518777174. <https://doi.org/10.1177/2331216518777174>
- Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology*, 51(3), 277–284.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear and Hearing*, 31(4), 480–490.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, 32(4), 498–510.
- Zhuang, J., Randall, B., Stamatakis, E. A., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The interaction of lexical semantics and cohort competition in spoken word recognition: An fMRI study. *Journal of Cognitive Neuroscience*, 23(12), 3778–3790.
- Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., & Marslen-Wilson, W. D. (2014). Optimally efficient neural systems for processing spoken language. *Cerebral Cortex*, 24(4), 908–918.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.