

A Tipping Point in Listening Effort: Effects of Linguistic Complexity and Age-Related Hearing Loss on Sentence Comprehension

Trends in Hearing
Volume 22: 1–14
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216518790907
journals.sagepub.com/home/tia



Nicole D. Ayasse¹  and Arthur Wingfield¹

Abstract

In recent years, there has been a growing interest in the relationship between effort and performance. Early formulations implied that, as the challenge of a task increases, individuals will exert more effort, with resultant maintenance of stable performance. We report an experiment in which normal-hearing young adults, normal-hearing older adults, and older adults with age-related mild-to-moderate hearing loss were tested for comprehension of recorded sentences that varied the comprehension challenge in two ways. First, sentences were constructed that expressed their meaning either with a simpler subject-relative syntactic structure or a more computationally demanding object-relative structure. Second, for each sentence type, an adjectival phrase was inserted that created either a short or long gap in the sentence between the agent performing an action and the action being performed. The measurement of pupil dilation as an index of processing effort showed effort to increase with task difficulty until a difficulty tipping point was reached. Beyond this point, the measurement of pupil size revealed a commitment of effort by the two groups of older adults who failed to keep pace with task demands as evidenced by reduced comprehension accuracy. We take these pupillometry data as revealing a complex relationship between task difficulty, effort, and performance that might not otherwise appear from task performance alone.

Keywords

pupillometry, cognitive aging, age-related hearing loss, listening effort

Date received: 7 February 2018; revised: 27 June 2018; accepted: 3 July 2018

Introduction

Since Kahneman's (1973) seminal publication of *Attention and Effort*, there has been a continuing interest in the relationship between effort and performance on cognitive tasks. A central feature of Kahneman's position was the postulate of a limited pool of processing resources that must be shared among simultaneous or closely sequential tasks such that the more resources required for successful accomplishment of one task, the fewer resources will be available for other tasks. Kahneman, who defined *effort* in terms of the allocation of processing resources, described a system in which resources may be allocated flexibly among operations according to their relative resource demands for their successful performance, and individual decisions as to task priorities.

It is Kahneman's limited-resource principle that underlies current arguments that, with degraded hearing, the perceptual effort needed for successful speech recognition draws resources that would otherwise be available

for encoding what has been heard in memory (Pichora-Fuller, Schneider, & Daneman, 1995; Rabbitt, 1968, 1991; Surprenant, 2007; Wingfield, Tun, & McCoy, 2005) and for comprehension of sentences that express their meaning with complex syntax (Wingfield, McCoy, Peelle, Tun, & Cox, 2006; see also Van Engen & Peelle, 2014).

One of the earliest extensions of Kahneman's emphasis on a relationship between effort and performance was seen in Norman and Bobrow's (1975) contrast between *resource-limited* processes, in which the degree of success is limited only by the amount of resources (effort) one is willing to devote to a task, and *data-limited* processes,

¹Department of Psychology and Volen National Center for Complex Systems, Brandeis University, Waltham, MA, USA

Corresponding Author:

Arthur Wingfield, Department of Psychology and Volen National Center for Complex Systems, Brandeis University, Waltham, MA, USA.
Email: wingfield@brandeis.edu



in which performance is limited by the quality of the stimulus input, such that no additional allocation of cognitive resources will further improve the level of performance (Norman & Bobrow, 1975, 1976; see also Kantowitz & Knight, 1976). Both of these notions, and Kahneman's limited-resource argument, have been embodied in the recent Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016) that has applied these notions more specifically to the challenges for speech comprehension imposed by hearing impairment.

Consistent with the earlier observations, FUEL defines effort as the allocation of mental resources to meet a perceptual or cognitive challenge, with the recognition that the effort put into a task will reflect a balance between listening difficulty, task demands, and motivation to expend the necessary effort to meet the processing challenge (Pichora-Fuller et al., 2016). One aspect of this relationship is the postulate that the degree of effort an individual is willing to expend toward a task will be greater if he or she believes that task success is possible, and hence that the required effort is justified (Richter, 2016).

The issue of processing effort takes on special significance in adult aging, where age-related hearing impairment is often accompanied by reduced working memory capacity and executive function (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Salthouse, 1994), and a general slowing in a number of perceptual and cognitive tasks (Salthouse, 1996). In this regard, it has been shown that older adults who can perform at ceiling or near-ceiling levels in comprehension of spoken sentences that have a simple syntactic structure (e.g., a canonical noun-verb-noun structure) may show significant difficulty comprehending sentences that express their meaning with complex syntax. Notably, this can occur even when the complex sentences contain exactly the same words as sentences with a simpler structure, are recorded by the same speaker, and are presented at the same sound level as the structurally simpler sentences (e.g., DeCaro, Peelle, Grossman, & Wingfield, 2016; Obler, Fein, Nicholas, & Albert, 1991; Stewart & Wingfield, 2009; Wingfield et al., 2006). Such findings, however, should not necessarily imply that older adults' successful comprehension of syntactically simpler sentences may not require more effort than young adults' to achieve the same level of comprehension performance. Nor do they directly address the question of whether listeners will continue to commit maximum effort when the linguistic complexity of the speech materials crosses a threshold of difficulty where successful comprehension may seem beyond reach.

In the following experiment, older adults with clinically normal hearing, age-matched older adults with mild-to-moderate hearing loss, and normal-hearing young

adults were tested for comprehension accuracy and processing effort for spoken English sentences that varied the comprehension challenge in two ways: increasing the syntactic complexity of a sentence and increasing the gap distance between the agent performing an action in a sentence and the action being performed.

Syntactic Manipulation

The simpler of two syntactic structures we used were sentences with a *subject-relative* embedded clause structure, such as "Sisters that assist brothers are generous." Subject-relative sentences follow a common pattern in English, in which the first noun in the sentence indicates an agent performing an action, the first verb identifies that action, and the second noun names the recipient of the action. Sentences with this structure were contrasted with sentences having the same words but with the meaning expressed with a more complex *object-relative* embedded clause structure, such as "Brothers that sisters assist are generous."

A number of reasons underlie the greater processing demands of sentences with an object-relative structure compared with those with a subject-relative structure. In object-relative sentences, the order of thematic roles is not canonical (i.e., the first noun is not the agent of the action) such that they require a more extensive thematic integration than subject-relative sentences (Warren & Gibson, 2002) and place a heavier demand on working memory (Cooke et al., 2002; DeCaro et al., 2016). In addition, because they are less common in everyday discourse (Goldman-Eisler, 1968; Goldman-Eisler & Cohen, 1970), they violate listener's expectations of the likely structure of a sentence being heard, thus requiring a reanalysis of the sentence meaning (cf., Gibson, Bergen, & Piantadosi, 2013; Levy, 2008; Novick, Trueswell, & Thompson-Schill, 2005; Padó, Crocker, & Keller, 2009). Consistent with these arguments, past studies have reliably shown that object-relative sentences produce more comprehension errors than subject-relative sentences, especially for older adults (e.g., Carpenter, Miyake, & Just, 1994; Wingfield, Peelle, & Grossman, 2003). For this reason, the subject-relative and object-relative contrast offers a good basis for an examination of performance and effort.

Agent-Action Gap Distance

To further vary the processing challenge, both the subject-relative and object-relative sentences we used had a four-word adjectival phrase inserted in a position that put a short or long gap between the agent performing an action and the action being performed. For example, the following is a subject-relative sentence with a short gap: "Sisters that assist brothers *with short brown hair* are generous." A sentence with the same subject-relative

structure but with a long gap between the agent and the action might be as follows: “Sisters *with short brown hair* that assist brothers are generous.” The following are the examples of sentences with a more complex object-relative structure with, respectively, short and long gaps: “Brothers *with short brown hair* that sisters assist are generous” and “Brothers that sisters *with short brown hair* assist are generous.” (In these examples, the agents and actions are underlined, and the inserted adjectival phrases are in italics.)

DeCaro et al. (2016) presented normal-hearing young adults and normal-hearing and hearing-impaired older adults with such sentences with instructions to indicate who was agent of the action. They found that all three groups’ comprehension was close to error-free for subject-relative sentences with a short gap between the agent and action. By contrast, all three participant groups began to make comprehension errors when hearing object-relative long gap sentences, with this degree of difficulty larger for the normal-hearing older adults than the young adults, and still larger for the hearing-impaired older adults (DeCaro et al., 2016). Consistent with a limited-resource model (Kahneman, 1973; Pichora-Fuller, 2016; see also Rabbitt, 1991; Wingfield et al., 2005), it was concluded that the listening effort associated with age-related hearing loss drew on older adults’ limited resources which, in turn, put a special burden on successful comprehension of the most computationally demanding object-relative long gap sentences. In offering this interpretation, however, DeCaro et al. acknowledged that the reference to effort was inferred from participants’ performance, rather than independently measured.

Measuring Processing Effort

A number of approaches have been taken to the measurement of processing effort and resource allocation associated with task performance. Dual-task studies have seen frequent use following the principle that the effort needed for success on, for example, a primary speech task will be revealed in a performance decrement on a concurrent unrelated task (Naveh-Benjamin, Craik, Guez, & Kreuger, 2005; Pals, Sarampalis, & Baskent, 2013; Sarampalis, 2009; Tun, Mccoy, & Wingfield, 2009). Such dual-task studies, however, are prone to trade-offs in the momentary attention given to each task that may complicate interpretation (Hegarty, Sha, & Miyake, 2000). Ratings of subjective effort, although having potential ecological interest, have shown mixed reliability as well as being an inherently off-line measure (see the review in McGarrigle et al., 2014). In contrast, *pupillometry* (the measurement of changes in the size of the pupil of the eye) can serve as an objective, online index of effort that can be measured independently of task performance.

Besides its response to ambient light and emotional arousal (Kim, Beversdorf, & Heilman, 2000), a transient change in pupil size has been shown to correspond to perceptual and cognitive effort (Beatty & Lucero-Wagoner, 2000; Kahneman & Beatty, 1966). Relevant to our present interests, numerous studies have reliably shown that increasing the perceptual or linguistic processing challenge for speech materials is accompanied by a progressive increase in pupil dilation. In the former case, this is true whether perceptual difficulty is increased by acoustically degrading the speech signal (e.g., Kramer, Kapteyn, Festen, & Kuik, 1997; Kuchinsky et al., 2013, 2014; Winn, 2016; Winn, Edwards, & Litovsky, 2015; Zekveld, Festen, & Kramer, 2013; Zekveld & Kramer, 2014; Zekveld, Kramer, & Festen, 2010, 2011) or whether the perceptual challenge results from impaired hearing acuity (e.g., Ayasse, Lash, & Wingfield, 2017; Kramer et al., 1997; Kuchinsky et al., 2013; Zekveld et al., 2011). Increased pupil dilation is also seen when listeners are tested for comprehension or recall of clearly spoken sentences that increase in linguistic complexity (e.g., Engelhardt, Ferreira, & Patsenko, 2010; Just & Carpenter, 1993; Piquado, Isaacowitz, et al., 2010; Wright & Kahneman, 1971).

Experimental Question

Testing comprehension of sentences that vary in syntactic complexity and gap distance, our experimental question was whether one would see a progressive increase in effort with increasing levels of sentence complexity, or whether older adults with hearing loss might reach a tipping point, in which the difficulty of the comprehension task is no longer accompanied by an increased commitment of effort.

The possibility that increasing task difficulty may lead to such a tipping point appeared early in the pupillometry literature. Such an effect was observed by Peavler (1974) in a digit-list recall study with young adults. Peavler observed that pupil dilation as an indication of cognitive effort was larger while recalling longer compared with shorter digit lists but that pupil dilation plateaued for supraspan lists (i.e., digit lists too long for accurate recall). These results suggest that, so long as additional effort results in additional gains, one will see the expected increase in pupil dilation as an index of this effort. Beyond this point, however, a further increase in task demands may result in a plateau, or potentially a decrease, in effort and its concomitant pupillary response.

Subsequent to Peavler’s (1974) findings for digit recall by young adults, analogous findings have appeared or been postulated in the context of comprehension of degraded speech and listeners’ willingness or ability to commit additional effort when difficulty exceeds a certain

point (cf., Kuchinsky et al., 2014; Ohlenforst et al., 2017; Richter, 2016; Wang et al., 2018; Wendt, Dau, & Hjortkjaer, 2016; Zekveld & Kramer, 2014; Zekveld et al., 2011). This question takes on special significance for older adults, and especially for those with hearing impairment, when faced with linguistically complex sentences as can occur in everyday speech communication.

By making continuous recordings of changes in pupil dilation while normal-hearing young adults, normal-hearing older adults, and hearing-impaired older adults were tested for sentence comprehension, we wished to test the alternative possibilities that (a) mean pupil sizes would be incrementally larger with increasing sentence complexity indicative of a progressive increase in commitment of effort, more so for older adults and older adults with impaired hearing or (b) whether pupil size as a measure of effort might plateau when sentence complexity reaches a difficulty threshold, with this potentially most likely to occur for older adults with impaired hearing.

Methods

Participants

Participants were 28 older adults, 14 with clinically normal hearing (3 males and 11 females) and 14 older adults with a mild-to-moderate hearing loss (4 males and 10 females). Audiometric assessment was conducted using an AudioStar Pro clinical audiometer (Grason-Stadler, Madison, WI, USA) using standard audiometric procedures in a sound attenuating testing room. The participants in the older adult normal-hearing group had a mean better ear pure tone average (PTA) of 18.75 dB HL ($SD = 3.98$) averaged across .5, 1, 2, and 4 kHz, and a mean better ear speech reception threshold (SRT) of 20.00 dB HL ($SD = 7.07$), placing them in a range typically defined in the audiological literature as clinically normal hearing for speech (PTA < 25 dB HL; Katz, 2002). The term *normal* is used here without the implication that this group's thresholds necessarily matched those of young adults with age-normal hearing acuity.

The older adult hearing-impaired group had a mean better ear PTA of 36.43 dB HL ($SD = 6.70$), and a mean better ear SRT of 33.93 dB HL ($SD = 8.13$), placing them in the mild-to-moderate hearing loss range (Katz, 2002). This degree of loss represents the most common degree of hearing loss among hearing-impaired older adults (Morrell et al., 1996). All participants had symmetrical hearing defined as an interaural PTA difference less than or equal to 15 dB HL. None of the participants reported regular use of hearing aids (e.g., Fischer et al., 2011), and all testing was conducted unaided.

The normal-hearing older adult group ranged in age from 69 to 79 years ($M = 72.90$ years, $SD = 3.13$), and

the hearing-impaired older adult group ranged in age from 68 to 85 years, $M = 74.47$ years, $SD = 4.65$; $t(26) = 1.05$, $p = .304$. The two groups were similar in verbal ability, as estimated by the Shipley vocabulary test (Zachary, 1991). This is a written multiple choice test in which the participant is required to indicate which of four listed words means the same or nearly the same as a given target word; M normal hearing = 16.93, $SD = 2.79$; M hearing impaired = 16.79, $SD = 1.97$; $t(26) = 0.16$, $p = .877$.

To ensure that the two older adult groups did not accidentally differ in cognitive ability, working memory capacity was assessed with the Reading Span task (RSpan) modified from Daneman and Carpenter (1980; Stine & Hindman, 1994). The RSpan task requires participants to read sets of sentences and respond after each sentence whether the statement in the sentence was true or false. Once a full set of sentences has been presented, participants are asked to recall the last word of each of the sentences in the order in which the sentences had been presented. Participants received three trials for any given number of sentences, with a working memory score calculated as the total number of trials in which all sentence-final words were recalled correctly in the correct order.

The RSpan task was chosen because it draws heavily on both storage and processing components that represent the characteristics of working memory (McCabe et al., 2010; Wingfield, 2016) and in written form would not be confounded with hearing acuity. Spans for the normal-hearing and hearing-impaired older adults, respectively, were 8.50 ($SD = 3.00$) and 7.43 ($SD = 2.38$), $t(26) = 1.05$, $p = .305$.

For purposes of comparison, we also included a group of 14 young adults (3 males, 11 females), ranging in age from 18 to 24 years ($M = 20.96$ years, $SD = 1.88$), all of whom had age-normal hearing acuity, with a mean PTA of 5.98 dB HL ($SD = 3.03$) and a mean SRT of 8.57 dB HL ($SD = 3.06$).

It is common for older adults to have superior vocabulary scores compared with young adults (e.g., Kempler & Zelinski, 1994; Verhaeghen, 2003). This held true for the young adults' vocabulary scores in the present sample, ($M = 13.64$, $SD = 1.78$) relative to both the normal-hearing, $t(26) = 3.72$, $p = .001$, and hearing-impaired, $t(26) = 4.43$, $p < .001$, older adults. As would be expected, the young adults' working memory RSpans ($M = 10.21$, $SD = 3.67$) tended to be larger than those of the older adults. This appeared as a nonsignificant trend when compared with the normal-hearing older adults, $t(26) = 1.35$, $p = .187$, but did reach significance when compared with the hearing-impaired older adults, $t(26) = 2.39$, $p = .025$.

All participants reported themselves to be in good health, with no history of stroke, Parkinson's disease,

or other neuropathology that might compromise their ability to carry out the experimental task. All participants reported themselves to be monolingual native speakers of American English. Written informed consent was obtained from all participants according to a protocol approved by the Brandeis University Institutional Review Board.

Stimuli

The stimuli were based on those used by DeCaro et al. (2016; see also Cooke et al., 2002). Stimulus preparation began with construction of 144 6-word sentences with a subject-relative structure that contained the agent of an action, the action, and the recipient of the action. In all cases, the agent and recipient could be plausibly interchanged. For each of these sentences, a counterpart sentence was constructed using the same six words but with the meaning expressed with an object-relative structure. Each of the subject-relative and object objective-relative sentences then had a four-word adjectival phrase inserted within the sentence either with at most a one-word separation between the agent and the action (*short gap sentences*) or in a position that separated the agent from the action by at least four words (*long gap sentences*). This resulted in a total of 576 10-word sentences (144 subject-relative short gap, 144 subject-relative long gap, 144 object-relative short gap, and 144 object-relative long gap). Examples of the four sentence types are shown in Table 1.

To discourage listeners from developing incidental processing strategies based in limited sentence types, 100 filler sentences were prepared in addition to the 144 test sentences. Fifty-two of these were 6- to

10-word sentences similar in content to the test sentences but that did not contain an embedded clause, and 48 consisted of 6-word sentences similar in structure to the test sentences but without the inclusion of a 4-word adjectival phrase.

The test sentences and fillers were recorded by a female native speaker of American English using Sound Studio v2.2.4 (Macromedia, Inc., San Francisco, CA, USA) that digitized (16-bit) at a sampling rate of 44.1 kHz. Recordings were equalized within and across sentence types for root-mean-square intensity using MATLAB (MathWorks, Natick, MA, USA).

Procedure

Stimulus presentations. Each participant heard 96 test sentences, 24 in each of the four sentence types (24 subject-relative short gap, 24 subject-relative long gap, 24 object-relative short gap, and 24 object-relative long gap) along with 100 filler sentences. Prior to each sentence presentation, the names of the agent and recipient of the action (e.g., *sisters*, *brothers*) were displayed horizontally on a computer screen. Participants were instructed to use the computer mouse to click on the correct name of the agent of the action in the just-heard sentence. The four sentence types and filler sentences were intermixed in presentation. Each core sentence (a particular combination of agent, action, and recipient) was heard only once by any participant, with the version in which each core sentence was heard counterbalanced across participants.

Stimuli were presented binaurally over Eartone 3A insert earphones (E-A-R Auditory Systems, Aero Company, Indianapolis, IN, USA) with a nominal presentation level of 20 dB above the individual's better ear SRT (20 dB SL). Prior to the main experiment, audibility was tested by presenting two low predictability sentences taken from the IEEE/Harvard sentence corpus (IEEE, 1969) at 20 dB SL with the instruction to repeat each sentence as it was presented. An example sentence was "The lake sparkled in the red, hot sun." Eleven older adults (7 normal hearing and 4 hearing impaired) were unable to accurately repeat back either of the two sentences. For these participants, the presentation level was increased to 25 dB SL, and two additional IEEE sentences were presented. All 11 were able to accurately repeat the sentences at this level. Twenty-five dB SL was used as the presentation level for these participants.

The main experiment was preceded by a brief practice session to familiarize the participant with the sound of the speaker's voice and the experimental procedures. Ten sentences, representing a mix of test sentence and filler types, were used in the practice session. None of these sentences was used in the main experiment.

Table 1. Examples of Sentence Types.

Sentence type	Agent-action distance	Example sentences
Subject-relative	Short gap	<u>Sisters</u> that <u>assist</u> brothers with short brown hair are generous
	Long gap	<u>Sisters</u> with short brown hair that <u>assist</u> brothers are generous
Object-relative	Short gap	Brothers with short brown hair that <u>sisters</u> <u>assist</u> are generous
	Long gap	Brothers that <u>sisters</u> with short brown hair <u>assist</u> are generous

Note. Underlined words indicate the agent performing the action and the action being performed.

Pupillary response data acquisition and preprocessing.

Throughout the course of each trial, the participant's moment-to-moment pupil size was recorded via an EyeLink 1000 Plus eye-tracking apparatus (SR Research, Ontario, Canada), with pupil size data acquired at a rate of 1000 Hz and recorded via MATLAB software (MathWorks, Natick, MA, USA). The EyeLink camera was positioned below the computer screen that showed the names of the agent and recipient for the particular sentence. To facilitate reliable pupil size measurement, the participant's head was stabilized using a customized individually adjusted chin rest that positioned the participant's eyes approximately 60 cm from the EyeLink camera.

Pupil diameters below three standard deviations of a trial mean were coded as a blink (Wendt et al., 2016; Zekveld et al., 2010, 2014; Zekveld & Kramer, 2014). These blinks were removed, and linear interpolation was performed starting 80 ms before and ending 160 ms after each blink. This procedure was used to reduce artifacts resulting from partial closures of the eyelids at the beginning and ending of a blink that would cause brief partial obscurations of the pupil (Siegle, Ichikawa, & Steinhauer, 2008; Winn et al., 2015). A 20-sample moving average smoothing filter was then passed over the data (e.g., Winn et al., 2015).

To adjust for individual differences in pupil size dynamic range, at the beginning of the session, pupil sizes were recorded while the participant viewed a light screen (199.8 cd/m²) presented for 60 s followed by a dark screen for 60 s (0.4 cd/m²). This range was used for calculation of adjusted pupil size as will be discussed. Ambient light in the testing room was kept constant throughout the experiment.

Peak pupil dilation (PPD) was quantified as the peak pupil size for correct trials occurring after the onset of the verb in the embedded clause and before the participant's response. Pupil size was baseline-corrected by subtracting measured pupil size from a pretrial baseline averaged over a 2-s window prior to the onset of the sentence. Pupil size changes were also scaled to account for age differences in the pupillary response (*senile miosis*; Bitsios, Prettyman, & Szabadi, 1996) by representing pupil size as a percentage ratio of the individuals' minimum constriction and maximum dilation in response to the light and dark screens presented. This was calculated as, $(d_M - d_{base}) / (d_{max} - d_{min}) \times 100$, where d_M is the participant's measured pupil size at a given time point, d_{base} is the participant's pretrial baseline, d_{min} is the participant's minimum constriction taken as the average pupil size over the last 30 s of viewing a light screen, and d_{max} is the participant's maximum dilation measured as the pupil size averaged over the last 30 s of viewing a dark screen (e.g., Allard, Wadlinger, & Isaacowitz, 2010; Ayasse et al., 2017; Piquado, Isaacowitz, et al., 2010).

Results

Comprehension Accuracy

Figure 1 depicts the accuracy data for all three participant groups. It can be seen that all three groups reached ceiling or near-ceiling accuracy levels for both the short and long gap subject-relative sentences. All three groups, however, appear to show poorer comprehension for object-relative compared with subject-relative sentences, and for the object-relative sentences, all three participant groups appear to perform worse in the long gap compared with the short gap condition. In addition, the presence of a long agent-action gap in the object-relative sentences appears to be associated with differentially poorer comprehension for the hearing-impaired older adults.

These data were analyzed using a logistic mixed-effects model, with syntax, gap distance, and participant group as fixed effects, and whether a trial was correct or incorrect as the dependent variable; individual participants and items were included as random effects using an intercept model based on Matuschek, Kliegl, Vasishth, Baayen, and Bates's (2017) suggested method for choosing the most parsimonious model. The fixed effects were added into the model in the aforementioned order with the respective interactions entered after the main effects. The effects of the fixed effects on model fit were evaluated using model comparisons of the change in log-likelihood using the analysis of variance function (Bates, Maechler, Bolker, & Walker, 2015; Matuschek et al., 2017), with the young adult group treated as the baseline for comparison. All analyses were carried out in

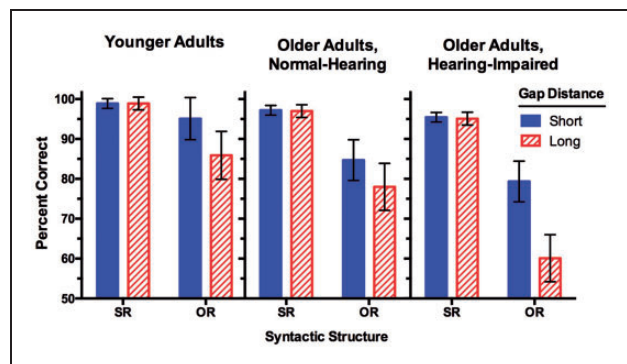


Figure 1. Mean comprehension accuracy for SR and OR sentences with a short or long gap distance between the agent performing an action and the action being performed. Data are shown for young adults with normal hearing acuity (left panel), older adults with clinically normal hearing for speech (middle panel), and older adults with a mild-to-moderate hearing loss (right panel). Error bars represent one standard error. OR = object-relative; SR = subject-relative.

R version 3.4.4 using the *lme4* package (version 1.1–15) and the function *glmer* to fit the models.

This analysis revealed a significant main effect of syntax, $B = -1.13$, $\chi^2(1) = 287.68$, $p < .001$, confirming that participants were significantly less accurate for object-relative compared with subject-relative sentences. There was also a significant main effect of gap distance, $B = -0.40$, $\chi^2(1) = 39.96$, $p < .001$, confirming that participants were overall less accurate for long compared with short gap sentences, as well as a significant Gap Distance \times Syntax interaction, $B = -0.26$, $\chi^2(1) = 10.89$, $p \leq .001$, driven by the larger effect of gap distance on accuracy in the object-relative compared with the subject-relative condition. In addition, there was a significant main effect of participant group, $B_{\text{OANH}} = -0.34$, $B_{\text{OAH}} = -0.38$, $\chi^2(2) = 7.12$, $p = .028$, driven primarily by the overall difference between the young adults and the hearing-impaired older adults, and a significant Participant Group \times Syntax interaction, $\chi^2(2) = 9.91$, $p = .007$, driven by the differential effect of more complex syntax on both of the older adult groups compared with the young adult group ($B_{\text{OANH}} = -0.27$, $B_{\text{OAH}} = -0.12$). The Participant Group \times Gap Distance interaction was not significant, $\chi^2(2) = 3.51$, $p = .173$.

A second analysis was conducted to compare the two groups of particular interest, the normal-hearing and hearing-impaired older adults (with the normal-hearing group treated as the baseline for comparison). This analysis was conducted on the object-relative sentences alone because participants from all groups reached ceiling or near-ceiling accuracy for the subject-relative sentences. A logistic mixed-effects model was run with gap distance and participant group as fixed effects, added in that order, and with the interaction term added last. Individual participants and items were entered as random effects. This analysis revealed a significant main effect of gap distance, $B = -0.56$, $\chi^2(1) = 43.38$, $p < .001$, again confirming that the older adult participants were less accurate for object-relative compared with subject-relative sentences. Although there was no main effect of participant group, $\chi^2(1) = 1.84$, $p = .175$, this was moderated by a significant Participant Group \times Gap Distance interaction, $B = -0.20$, $\chi^2(1) = 5.11$, $p = .024$. This interaction reflects the larger effect of gap distance experienced by the hearing-impaired older adults compared with the normal-hearing older adults.

Pupillary Responses

Figure 2 shows the mean-adjusted PPD associated with correct comprehension of subject-relative and object-relative short and long gap sentences for each of the three participant groups. Two important features are suggested by visual inspection of Figure 2. First, there appears a progressive increase in PPD across the

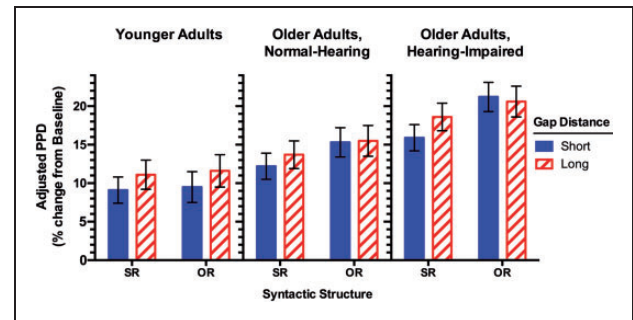


Figure 2. Mean-adjusted peak pupil size associated with comprehension of SR and OR sentences with a short or long gap distance between the agent performing an action and the action being performed. Data are shown for young adults with normal hearing acuity (left panel), older adults with clinically normal hearing for speech (middle panel), and older adults with a mild-to-moderate hearing loss (right panel). Error bars represent one standard error. OR = object-relative; PPD = peak pupil dilation; SR = subject-relative.

three participant groups, with the older adult hearing-impaired group showing the largest PPD and the young adults the smallest. Second, while the young adults show an increase in PPD in response to long gap sentences relative to short gap sentences for both subject-relative and object-relative sentences, the two older adult groups show this only for the subject-relative sentences. For the more challenging object-relative sentences, adding a long gap between agent and action in the sentences did not result in a further increase in PPD for the older adults.

These data were analyzed using a linear mixed-effects model in a similar manner to the comprehension accuracy data but the use of the *lmer* function and the adjusted PPD as the dependent variable. This analysis revealed a significant main effect of syntax, $B = 1.04$, $\chi^2(1) = 30.50$, $p < .001$, confirming that participants exhibited an overall larger PPD for the object-relative compared with the subject-relative sentences. There was also a significant main effect of gap distance, $B = 0.48$, $\chi^2(1) = 6.89$, $p = .009$, confirming that participants also exhibited an overall larger PPD for the long compared with the short gap sentences. There was a significant main effect of participant group, $B_{\text{OANH}} = 1.64$, $B_{\text{OAH}} = 2.27$, $\chi^2(2) = 8.73$, $p = .013$, driven primarily by the difference between the young adults and the hearing-impaired older adults ($p = .003$), as well as a significant Participant Group \times Syntax interaction, $B_{\text{OANH}} = 0.48$, $B_{\text{OAH}} = 0.31$, $\chi^2(2) = 9.36$, $p = .009$, confirming that the older adult groups exhibited a larger effect of more complex syntax on PPD than the young adults. The Syntax \times Gap Distance interaction was not significant, $\chi^2(1) = 2.83$, $p = .093$, nor was there a Participant Group \times Gap Distance interaction, $\chi^2(2) = .06$, $p = .971$.

A second analysis was conducted to compare the two groups of primary interest, the normal-hearing and hearing-impaired older adults. Again, syntax, gap distance, and group were entered into a linear mixed model in that order with the respective interactions entered after the main effects. Participants and items were included as random effects and normal-hearing older adults were treated as the baseline for purposes of comparison. This analysis revealed a significant main effect of syntax, $B=1.39$, $\chi^2(1)=38.98$, $p<.001$, confirming that older adult participants showed a larger PPD for object-relative than for subject-relative sentences. There was a significant main effect of gap distance, $B=0.49$, $\chi^2(1)=5.22$, $p=.022$, confirming that participants showed a larger PPD for long gap than for short gap sentences, and a significant Gap Distance \times Syntax interaction, $B=-0.45$, $\chi^2(1)=4.32$, $p=.038$, driven by the larger effect on PPD of gap distance in the subject-relative condition than in the object-relative condition for the these two participant groups. There was a marginal main effect of group, $\chi^2(1)=3.23$, $p=.072$. Neither the Gap Distance \times Participant Group, $\chi^2(1)=.08$, $p=.784$, nor the Syntax \times Participant Group, $\chi^2(1)=.99$, $p=.320$, interactions were significant.

An alternative to peak pupil diameter as an index of effort is the calculation of mean pupil size across a selected region of a trial (cf., Ahern & Beatty, 1979; Verney, Granholm, & Dionisio, 2001; Zekveld et al., 2010). For comparison, we calculated the mean pupil size over a 2s time bin beginning at the onset of the verb in the embedded clause, the approximate time point at which the meaning of the sentence could be resolved. Two seconds was chosen as a sufficient window to capture the majority of the pupillary response to cognitive processing (Bitsios et al., 1996). At least for our data, calculation of mean pupil size yielded a similar pattern in response to sentence type and participant group as observed with PPD.

Effects of Working Memory Capacity and Hearing Acuity as Continuous Variables

As would be expected from extant literature, the young and older adults in this study differed in both working memory capacity and hearing acuity. This raises the question of the degree to which each of these variables may have contributed to the group differences in comprehension accuracy and processing effort as indexed by the pupillary response. To address this question, we conducted mixed-model analyses using the continuous variables of hearing acuity, working memory, and age as predictors, first of comprehension accuracy, and then the size of the pupillary response.

Predictors of comprehension accuracy. Because one or more of the predictor variables might operate differently

depending on the linguistic-level challenge imposed by the sentence types, separate analyses were conducted for the subject-relative and object-relative short and long gap sentences. Logistic mixed-effects models were run on each experimental condition, with whether the participant was correct or incorrect for any given trial serving as the dependent variable in each case, and individual participants and items included as random effects. The three predictor variables were entered into the model in the following order: hearing acuity (represented by the better ear PTA, averaged over .5, 1, 2, and 4 kHz), working memory span (represented by RSpan), and participants' chronological age in years. The effects of each variable on model fit were evaluated using model comparisons of the change in log-likelihood, and all analyses were carried out in R version 3.4.4 using the *lme4* package (version 1.1–15) and the *glmer* function. The predictor variables were rescaled and centered according to their standard deviations before conducting the analyses. Table 2 shows the coefficients and χ^2 test results for each step of the model.

For the subject-relative short and long gap sentences, there was no significant main effect of adding hearing acuity, working memory, or age onto the model. However, for the object-relative short gap sentences, there was a significant effect of adding hearing acuity into the model, while the effects of working memory

Table 2. Logistic Mixed-Effects Models of Continuous Variables for Comprehension Accuracy.

	Predictor	B^a	χ^2^b	df^c	p^d
SR Short ^e	Hearing acuity	-0.184	0.52	1	.469
	Reading Span	-0.002	<0.01	1	.995
	Age	-0.455	1.15	1	.284
SR Long	Hearing acuity	-0.367	1.66	1	.197
	Reading Span	0.241	0.54	1	.463
	Age	-0.274	0.34	1	.562
OR Short	Hearing acuity	-0.836	6.10	1	.014*
	Reading Span	0.003	<0.01	1	.994
	Age	-0.458	0.78	1	.376
OR Long	Hearing acuity	-0.766	8.87	1	.003**
	Reading Span	0.541	4.31	1	.038*
	Age	-0.028	<0.01	1	>.999

Note. Significant p values are bolded.

^aUnstandardized coefficient (of standardized variables).

^b χ^2 value for comparisons of each step of the model.

^cDegrees of freedom for the χ^2 test.

^d p value reflects significance of change in model fit at each step of the model.

^eSR and OR indicate subject-relative and object-relative syntactic constructions, respectively. The terms *Long* and *Short* refer to a short or long gap distance between the agent and action in a sentence.

* $p < .05$. ** $p < .01$.

and age were not significant once hearing acuity had been accounted for. For the most challenging condition, the object-relative long gap sentences, both hearing acuity and working memory, were significant predictors, while age again was not. That is, increased hearing thresholds (decreased hearing acuity) predicted decreased accuracy, and increased working memory capacity predicted increased accuracy.

The absence of a contribution of any of the predictor variables for the subject-relative sentences in either their short gap or long gap versions can be attributed to the participants' near-ceiling performance for these less challenging sentence types. In the case of the object-relative short gap sentences, one sees a significant contribution of hearing acuity, indicative of a single-resource model in which effortful listening attendant to hearing impairment had a detrimental effect on comprehension accuracy even though presentations were at a perceptually audible level. However, when a long gap was imposed between the agent and action in the already challenging object-relative sentences, a significant contribution of hearing acuity was joined by a significant effect of working memory span as predictors of comprehension accuracy. Once these sensory and cognitive variables were taken into account, participant age added no additional significant effect.

Predictors of peak pupillary response. A linear mixed-model analysis was run for the pupillary responses using the same predictor variables and order of entry as was used for comprehension accuracy, using the *lmer* function in the *lme4* package (version 1.1–15). Table 3 shows the coefficients and χ^2 test results for each step of the model. In all four conditions, hearing acuity was a significant predictor of PPD such that increased hearing thresholds (decreased hearing acuity) predicted a larger PPD. Neither working memory nor age contributed any additional significant amount in any of the four conditions.

Discussion

The so-called *effortfulness hypothesis* was introduced in a series of articles by Rabbitt (1969, 1991; Dickinson & Rabbitt, 1991) following his demonstration that recall of degraded auditory or visual stimuli was depressed, even though he could show that the stimuli themselves had been correctly recognized. Rabbitt's account for this finding was a principle akin to current limited-resource formulations. That is, Rabbitt argued that the draw on cognitive resources needed for successful identification of degraded auditory (Rabbitt, 1968, 1991) or visual (Dickinson & Rabbitt, 1991) stimuli left fewer resources available for successfully completing subsequent memory encoding operations. We discuss this principle first in

Table 3. Linear Mixed-Effects Models of Continuous Variables for Peak Pupil Dilation.

	Predictor	B^a	χ^2^b	df^c	p^d
SR Short ^e	Hearing acuity	2.68	6.33	1	.012*
	Reading Span	0.08	0.03	1	.855
	Age	0.33	0.04	1	.837
SR Long	Hearing acuity	2.91	6.60	1	.010**
	Reading Span	−0.84	0.52	1	.473
	Age	1.57	0.79	1	.374
OR Short	Hearing acuity	3.07	5.00	1	.025*
	Reading Span	1.33	1.04	1	.307
	Age	1.14	0.315	1	.575
OR Long	Hearing acuity	3.73	6.71	1	.010**
	Reading Span	−0.58	0.17	1	.676
	Age	0.50	0.06	1	.814

Note. Significant p values are bolded.

^aUnstandardized coefficient (of standardized variables).

^b χ^2 value for comparisons of each step of the model.

^cDegrees of freedom for the χ^2 test.

^d p value reflects significance of change in model fit at each step of the model.

^eSR and OR indicate subject-relative and object-relative syntactic constructions, respectively. The terms *Long* and *Short* refer to a short or long gap distance between the agent and action in a sentence.

* $p < .05$. ** $p < .01$.

terms of comprehension accuracy, and then in terms of our data on the pupillary response.

Comprehension Accuracy

As would have been expected from prior research (DeCaro et al., 2016), both syntactic complexity and the gap between the agent performing an action in a sentence and the action being performed had significant effects on comprehension accuracy. The present finding that all three participant groups performed at ceiling or near-ceiling regardless of gap distance for subject-relative sentences is emblematic of older adults' generally effective comprehension of spoken sentences when meaning is expressed in a canonical syntactic form and presented at a suprathreshold intensity level. When the sentence meaning was expressed with a more complex object-relative structure, however, an age difference now appeared. Most striking was a sharp decline in comprehension accuracy for object-relative sentences with a long agent-action gap for the older adults with hearing impairment when compared with older adults with better hearing acuity.

In noting this differential effect of linguistic challenge on the hearing-impaired older adults' comprehension, it is important to emphasize that the long gap object-relative sentences had the same words, were recorded by the same speaker, and were presented at the same

sound level as the accurately comprehended subject-relative short and long gap sentences. Consistent with Rabbitt's (1968, 1991) effortfulness hypothesis, and the related central resource models (Kahneman, 1973; Pichora-Fuller et al., 2016; Wingfield, 2016), we would interpret this finding as demonstrating that the extra resources (effort) required by the hearing-impaired participants for success at the perceptual level drew on resources that would otherwise be available for higher level comprehension operations. This extra draw on resources would have little overt consequence for comprehension of computationally less demanding speech materials such as the subject-relative sentences used in the present experiment and might go unnoticed in everyday discourse. A detrimental effect of this same resource draw, however, would be revealed when the hearing-impaired listener is confronted by speech materials with high resource demands at the linguistic level, such as revealed in the present experiment with the object-relative long gap sentences.

A possible mechanism underlying Rabbitt's *effortfulness effect* with acoustically impoverished but identifiable stimuli may be a slowing of perceptual identification due to inferential or top-down operations being brought into play to supplement the impoverished bottom-up information (Rönnerberg et al., 2013; Wingfield et al., 2015). In the case of word sequences, such slowing to completion would result in perceptual operations being still ongoing on one word as the next word is arriving. The effect would be to interfere with encoding of the prior word, with such effects potentially cascading to inefficient encoding of the full word sequence.

Although this concept has been tested at the level of word-list recall (Cousins, Dar, Wingfield, & Miller, 2014; Miller & Wingfield, 2010; Piquado, Cousins et al., 2010), one may speculate that a similar principle of sensory-based interference underlies errors in comprehension of sentence meanings, with such an effect appearing most prominently for computationally demanding sentences such as those used in the present experiment. This postulate might imply that the increased effect of interference would also be revealed in longer latencies to peak pupillary responses. This possibility was examined but did not appear in the present study. However, slowed perceptual processing could interfere with comprehension at the sentence level without necessarily resulting in a delayed pupillary response. This would be so, for example, if the primary effect of the interference is to reduce available resources for additional poststimulus processing. We suggest this as an area for future research.

Whatever the mechanism underlying the interference effect of degraded but identifiable stimuli on comprehension and memory, the comprehension data in the present experiment add to a number of studies showing poorer

comprehension and recall of impoverished but supra-threshold auditory stimuli, and especially so for older adults, and older adults with mild-to-moderate hearing impairment (e.g., DeCaro et al., 2016; Pichora-Fuller et al., 1995; Surprenant, 2007; Ward et al., 2016; Wingfield et al., 2005, 2006; Winn, 2016).

Pupillary Response

Our pupillometry results join others in showing that the task-evoked pupillary response is sensitive to task difficulty. Such studies, some conducted with young adults and some with middle aged or older adults, have shown increased pupil dilatation when listeners have been presented with speech that has been acoustically degraded, that has complex syntax, or that has lacked helpful contextual constraints (e.g., Just & Carpenter, 1993; Kramer et al., 1997; Piquado, Isaacowitz, et al., 2010; Wendt et al., 2016; Winn, 2016; Winn et al., 2015; Zekveld et al., 2010, 2011; Zekveld & Kramer, 2014). Our present findings reveal the sensitivity of the pupillary response in a number of expected but previously untested ways. There were also findings that might have been less expected.

Among our expected outcomes, we observed that even though comprehension accuracy was at or near ceiling for the subject-relative sentences regardless of gap distance, comprehension of the long gap subject-relative sentences tended to be accompanied by larger pupil dilations than short gap subject-relative sentences across the three participant groups. As also might have been predicted from an extension of Rabbitt's (1968, 1991) effortfulness hypothesis, decreased hearing acuity predicted a larger pupillary response. The importance of hearing acuity to processing effort was further confirmed for all four sentence types when hearing acuity, age, and working memory were considered as continuous variables.

Treating hearing acuity, working memory as measured by the RSpan, and age as continuous variables showed hearing acuity to have a significant contribution to comprehension accuracy only for the more resource-demanding object-relative sentences, while hearing contributed significantly to pupillary responses for all four sentence types. Working memory appeared as a factor on comprehension only for the most computationally demanding object-relative long gap sentences and not at all on pupillary responses.

A potentially less intuitive finding was the absence of an increase in pupil dilation for the more challenging long gap object-relative sentences compared with the short gap object-relative sentences for the two older adult participant groups. To the extent that the pupillary response serves as an index of processing effort, this would appear to reflect a plateau in the amount of effort the older adults were able, or willing, to commit to the comprehension task when that task had reached a

tipping point of processing difficulty marked by a combination of complex syntax and a long agent-action gap. We consider this possibility in the following section.

Effort as Task Engagement

Although a direct extrapolation from the less complex sentence types might lead one to expect a further increase in task difficulty to be accompanied by a further increase in relative pupil dilation, we saw instead a more complex relationship between pupil size as an index of effort and processing challenge.

A plateau in pupil size when task difficulty begins to exceed processing capacity is not without precedence in the literature. Peavler (1974), for example, reported that pupil size plateaued when the size of digit lists exceeded young adults' digit spans, while Zekveld and Kramer (2014), in a speech in noise study, observed a plateau in pupil dilation and a decrease in self-reports of expended effort, at especially low intelligibility levels (see also Kuchinsky et al., 2014; Ohlenforst et al., 2017; Wang et al., 2018).

The question might thus be raised as to whether there is a specific level of difficulty that can define a difficulty tipping point. In a speech in noise recall task, Ohlenforst et al. (2017) found the pupillary response to peak at approximately 50% accuracy in that half of the sentences were recalled correctly. In the present study, we saw a higher accuracy level leading to a plateau in pupil dilation for the older adults. It is difficult to compare these two findings, however, as Ohlenforst et al. (2017) were testing recall, with scores that could vary from 0% to 100%, while in the present study, we tested comprehension, where simple chance would yield 50% correct.

We cannot say with our current data whether the plateau in effort we saw for the older adults in the most difficult linguistic condition (long gap object-relative sentences) reflects a reduced ability to engage in effortful processing consequent to age-related changes in frontal attention networks, or an unwillingness to expend the necessary effort when the task demands make it uncertain whether additional effort will achieve success (cf., Kuchinsky et al., 2014; Richter, 2016). In either case, it can be seen that the plateau of effortful engagement was associated with markedly reduced comprehension accuracy for the hearing-impaired older adults under the dual challenge of object-relative sentences with a long agent-action gap.

Conclusions

It is the case that *effort*, like *resources*, has yet to be fully defined (McGarrigle et al., 2014; Wingfield, 2016). Indeed, one often sees the terms *resources*, *processing resources*, and *attentional resources* used interchangeably in the cognitive literature. It must also be acknowledged

that, because of the generally smaller size and more limited dynamic range in pupil size in older adults (*senile miosis*; Bitsios et al., 1996), using pupil dilation to compare degrees of effort across age groups requires caution (cf., Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004; Wang et al., 2016).

Two points are nevertheless clear from the current study. The first is that differences in effort were observed even when comprehension accuracy was at near-ceiling levels of performance, demonstrating the importance of additional metrics for evaluating task difficulty beyond comprehension accuracy or intelligibility alone. The second is an observation consistent with a tipping point principle in performance and effort. That is, although ideally the effort given to a task should increase with the challenge represented by the task, in cases where a task crosses a threshold of difficulty task-evoked pupillary responses may reveal an inadequate commitment of effort, potentially associated with a decline in task success. It is thus clear that the use of pupillometry reveals a complex relationship between task difficulty, effort and performance that might not otherwise appear from task performance alone.

Acknowledgments

We thank Mario Svirsky for his insightful comments on a tipping point in task difficulty and its implications for performance. We also thank Victoria Sorrentino for her help in data collection.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institutes of Health under award R01 AG019714 from the National Institute on Aging (to A. W.). N. D. A. acknowledges support from NIH training grant T32 GM084907. The authors also gratefully acknowledge the support from the W.M. Keck Foundation.

ORCID iD

Nicole D Ayasse  <http://orcid.org/0000-0003-0890-7742>

References

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*, 1289–1292. doi: 10.1126/science.472746
- 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. *Aging, Neuropsychology, and Cognition*, *17*(3), 296–311. doi: 10.1080/13825580903265681

- Ayasse, N. D., Lash, A., & Wingfield, A. (2017). Effort not speed characterizes comprehension of spoken sentences by older adults with mild hearing impairment. *Frontiers in Aging Neuroscience*, *8*, 1–12. doi: 10.3389/fnagi.2016.00329
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds), *Handbook of psychophysiology* (pp. 142–162). Cambridge, England: Cambridge University Press.
- 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*, 432–438. doi: 10.1093/ageing/25.6.432
- Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075–1122). San Diego, CA: Academic Press.
- Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., ... Grossman, M. (2002). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping*, *15*(2), 80–94. doi: 10.1002/hbm.10006
- 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 & Cognition*, *42*(4), 622–638. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24838269>. doi: 10.3758/s13421-013-0377-7
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *23*, 569–578. doi: 10.1016/S0022-5371(80)90312-6
- DeCaro, R., Peelle, J. E., Grossman, M., & Wingfield, A. (2016). The two sides of sensory-cognitive interactions: Effects of age, hearing acuity, and working memory span on sentence comprehension. *Frontiers in Psychology*, *7*, 236. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/26973557>. doi: 10.3389/fpsyg.2016.00236
- Dickinson, C. V. M., & Rabbitt, P. M. A. (1991). Simulated visual impairment: Effects on text comprehension and reading speed. *Clinical Vision Sciences*, *6*, 301–308.
- Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*, 639–645. doi: 10.1080/17470210903469864
- Fischer, M. E., Cruickshanks, K. J., Wiley, T. L., Klein, B. E., Klein, R., & Tweed, T. S. (2011). Determinants of hearing aid acquisition in older adults. *American Journal of Public Health*, *101*, 1449–1455. doi: 10.2105/AJPH.2010.300078
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056. doi: 10.1037/pnas.1216438110
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York, NY: Academic Press.
- Goldman-Eisler, F., & Cohen, M. (1970). Is N, P, and NP difficulty a valid criterion of transformational operations? *Journal of Verbal Learning and Verbal Behavior*, *9*, 161–166.
- Hegarty, M., Sha, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, *28*(3), 376–385.
- IEEE. (1969). Subcommittee on Subjective Measurements IEEE Recommended Practices for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 227–246.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*, 310–339.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. doi: 10.1126/science.154.3756.1583
- Kantowitz, B. H., & Knight, J. L. (1976). On experimenter limited processes. *Psychological Review*, *83*, 502–507. doi: 10.1037/0033-295X.83.6.502
- Katz, J. (Ed.). (2002). *Handbook of clinical audiology* (5th ed.). Philadelphia, PA: Lippincott, Williams and Wilkins.
- Kempler, D., & Zelinski, E. M. (1994). Language in dementia and normal aging. In F. A. Hupper, C. Brayne, & D. W. O'Conner (Eds), *Dementia and normal aging* (pp. 331–364). Cambridge, England: Cambridge University Press.
- Kim, M., Beversdorf, D. Q., & Heilman, K. M. (2000). Arousal response with aging: Pupillographic study. *Journal of the International Neuropsychological Society*, *6*(3), 348–350.
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilation. *Audiology*, *36*, 155–164. doi: 10.3109/00206099709071969
- 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. *Psychophysiology*, *51*, 1046–1057. doi: 10.1111/psyp.12242
- 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. doi: 10.1111/j.1469-8986.2012.01477.x
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177. doi: 10.1016/j.cognition.2007.05.006
- Matuschek, H., Kliegl, R., Vasisht, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi: 10.1016/j.jml.2017.01.001
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. doi: 10.1037/a0017619
- McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014).

- Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, 53(7), 433–440. doi: 10.3109/14992027.2014.890296
- Miller, P., & Wingfield, A. (2010). Distinct effects of perceptual quality on auditory word recognition, memory formation and recall in a neural model of sequential memory. *Frontiers in Systems Neuroscience*, 4, 14. doi: 10.3389/fnsys.2010.00014
- Morrell, C. H., Gordon-Salant, S., Pearson, J. D., et al. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *Journal of the Acoustical Society of America*, 100(4): 1949–1967.
- Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 520–537. doi: 10.1037/0278-7393.31.3.520
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44–64.
- Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review*, 83, 508–510. doi: 10.1037/0033-295X.83.6.508
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Re-examining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5, 263–281.
- Obler, L. K., Fein, D., Nicholas, M., & Albert, M. L. (1991). Auditory comprehension and aging: Decline in syntactic processing. *Applied Psycholinguistics*, 12(4), 433–452. doi: 10.1017/S0142716400005865
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., ... Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, 351, 68–79. doi:10.1016/j.heares.2017.05.012
- Padó, U., Crocker, M. W., & Keller, F. (2009). A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science*, 33, 794–838. doi: 10.1111/j.1551-6709.2009.01033.x
- Pals, C., Sarampalis, A., & Baskent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research*, 56, 1075–1084. doi: 10.1044/1092-4388(2012/12-0074)
- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, 11, 559–566. doi: 10.1111/j.1469-8986.1974.tb01114.x
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, 37, 5S–27S. doi: 10.1097/AUD.0000000000000312
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97, 593–607. doi: 10.1121/1.412282
- Piquado, T., Cousins, K. A. Q., Wingfield, A., & Miller, P. (2010). Effects of degraded sensory input on memory for speech: Behavioral data and a test of biologically constrained computational models. *Brain Research*, 1365, 48–65. doi: 10.1016/j.brainres.2010.09.070
- Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47(3), 560–569. doi:10.1111/j.1469-8986.2009.00947.x
- Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, 20(3), 241–248. doi:10.1080/14640746808400158
- Rabbitt, P. M. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica*, 111(Sup 476), 167–176. doi:10.3109/00016489109127274
- Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, 37, 111S–117S. doi:10.1097/AUD.0000000000000295
- Rönnerberg, J., Lunner, T., Zekveld, A., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7(31). doi: 10.3389/fnsys.2013.00031
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535–543.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–428.
- Sarampalis, A. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, 52, 1230–1240. doi:10.1044/1092-4388(2009/08-0111)
- Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45, 679–687. doi: 10.1111/j.1469-8986.2008.00681.x
- Stewart, R., & Wingfield, A. (2009). Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. *Journal of the American Academy of Audiology*, 20(2), 147–154.
- Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging Cognition*, 1, 2–16. doi:10.1080/09289919408251446
- Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging, Neuropsychology, and Cognition*, 14, 126–143. doi:10.1080/13825580701217710
- Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, 24(3), 761–766. doi:10.1037/a0014802
- Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 1–4. Retrieved from <http://journal.frontiersin.org/article/10.3389/fnhum.2014.00577/abstract>. doi:10.3389/fnhum.2014.00577
- Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41(2), 167–174. doi:10.1111/j.1469-8986.2003.00148.x

- Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18*(2), 332–339. doi:10.1037/0882-7974.18.2.332
- Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology, 38*, 76–83.
- Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear and Hearing, 39*(3), 573–582. doi:10.1097/AUD.0000000000000512
- Wang, Y., Zekveld, A. A., Naylor, G., Ohlenforst, B., Jansma, E. P., Lorens, A., . . . Kramer, S. E. (2016). Parasympathetic nervous system dysfunction, as identified by pupil light reflex, and its possible connection to hearing impairment. *PLoS One, 11*(4), 1–26. doi:10.1371/journal.pone.0153566
- 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), 97–111. doi: 10.1080/0361073X.2016.1108785
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79–112. doi:10.1016/S0010-0277(02)00087-2
- Wendt, D., Dau, T., & Hjortkjaer, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology, 7*, 1–12. doi:10.3389/fpsyg.2016.00345
- Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing, 37*, 35S–43S. doi:10.1097/AUD.0000000000000310
- Wingfield, A., Amichetta, N. M., & Lash, A. (2015). Cognitive aging and hearing acuity: modeling spoken language comprehension. *Frontiers in Psychology, 6*(684). doi: 10.3389/fpsyg.2015.00684
- 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. doi:10.3766/jaaa.17.7.4
- Wingfield, A., Peelle, J. E., & Grossman, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Aging, Neuropsychology, and Cognition, 10*, 310–322.
- 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. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing, 20*, 1–17. doi: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), 1–13. doi:10.1097/AUD.0000000000000145
- Wright, P., & Kahneman, D. (1971). Evidence for alternative strategies of sentence retention. *Quarterly Journal of Experimental Psychology, 23*, 197–213.
- Zachary, R. A. (1991). *Shipley Institute of Living Scale: Revised manual*. Los Angeles, CA: Western Psychological Services.
- 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. doi:10.1111/psyp.12151
- 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*, 480–490. doi:10.1097/AUD.0b013e3181d4f251
- 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*, 498–510. doi: 10.1097/AUD.0b013e31820512bb
- Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2013). Task difficulty differentially affects two measures of processing load: The pupil response during sentence processing and delayed cued recall of the sentences. *Journal of Speech, Language, and Hearing Research, 56*, 1156–1165. doi:10.1044/1092-4388(2012/12-0058)
- Zekveld, A. A., Rudner, M., Kramer, S. E., et al. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience, 8*(88). doi: 10.3389/fnins.2014.00088