

# Pupillometry Reveals That Context Benefit in Speech Perception Can Be Disrupted by Later-Occurring Sounds, Especially in Listeners With Cochlear Implants

Trends in Hearing  
Volume 22: 1–22  
© The Author(s) 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2331216518808962  
journals.sagepub.com/home/tia



Matthew B. Winn<sup>1</sup> and Ashley N. Moore<sup>1</sup>

## Abstract

Contextual cues can be used to improve speech recognition, especially for people with hearing impairment. However, previous work has suggested that when the auditory signal is degraded, context might be used more slowly than when the signal is clear. This potentially puts the hearing-impaired listener in a dilemma of continuing to process the last sentence when the next sentence has already begun. This study measured the time course of the benefit of context using pupillary responses to high- and low-context sentences that were followed by silence or various auditory distractors (babble noise, ignored digits, or attended digits). Participants were listeners with cochlear implants or normal hearing using a 12-channel noise vocoder. Context-related differences in pupil dilation were greater for normal hearing than for cochlear implant listeners, even when scaled for differences in pupil reactivity. The benefit of context was systematically reduced for both groups by the presence of the later-occurring sounds, including virtually complete negation when sentences were followed by another attended utterance. These results challenge how we interpret the benefit of context in experiments that present just one utterance at a time. If a listener uses context to “repair” part of a sentence, and later-occurring auditory stimuli interfere with that repair process, the benefit of context might not survive outside the idealized laboratory or clinical environment. Elevated listening effort in hearing-impaired listeners might therefore result not just from poor auditory encoding but also inefficient use of context and prolonged processing of misperceived utterances competing with perception of incoming speech.

## Keywords

listening effort, cochlear implants, pupillometry, speech perception, context

Date received: 15 January 2018; revised: 26 September 2018; accepted: 1 October 2018

## Introduction

Listening effort is an important aspect of hearing loss (HL) that has received increased attention in recent years. Listening to speech is more effortful for individuals with HL compared with those with normal hearing (NH; Hicks & Tharpe, 2002; Kramer, Kapteyn, Festen, & Kuik, 1997). As a result, people with HL not only are less accurate at hearing speech, but they also must work harder to achieve that level of understanding. Elevated effort associated with HL has been linked to greater fatigue (Edwards, 2007) and increased levels of anxiety and feelings of loss of control at work (Morata et al., 2005). Individuals with HL who report increased listening effort also report taking more sick-leave from work due to

mental distress (Kramer, 2008; Kramer, Kapteyn, & Houtgast, 2006). Given that one in seven U.S. adults are affected by HL (Hoffman, Dobie, & Losonczy, 2017), listening effort could have enormous impact on society in numerous ways. It is essential to understand what factors increase effort, and what can be done to alleviate it. There are numerous aspects of effort that

<sup>1</sup>Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, USA

## Corresponding author:

Matthew B. Winn, Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, USA.

Email: mwinn@umn.edu



can be examined, ranging from low-level sensory encoding to higher level linguistic comprehension and prediction. In this study, we focus on a situation that likely has disproportionate impact on people with HL, where effort remains elevated because a listener is hearing a new sound before the previous sentence was fully understood.

Listening effort is not explicitly measured by most clinical tests of speech perception; two people with the same intelligibility score may require different levels of effort to achieve that score. The effort to understand words is therefore not indicated by the score itself. Multiple studies show that effort can change significantly—and for different reasons—even when intelligibility is held constant (Francis, MacPherson, Chandrasekaran, & Alvar, 2016; Koelewijn, Zekveld, Festen, & Kramer, 2012). Even when differences in intelligibility are magnified by the introduction of noise, it can be unclear how much the noise-impacted scores are driven by various factors like stream segregation, attention, or linguistic recovery of masked words (see Mattys, Brooks, & Cooke, 2009 and Francis, 2010, for examples, of how these effects could be disentangled). Intelligibility scores are a combination of real-time auditory encoding, linguistic processing, and follow-up cognitive repair that could correct mistakes in word recognition. The incremental nature of word recognition has been explored in great detail, suggesting that it involves multiple linguistic processes that act in parallel and which interact with the incoming acoustic signal (Alloppenna, Magnuson, & Tanenhaus, 1998; Luce & Pisoni, 1998; McClelland & Elman, 1986). These processes likely demand different kinds of cognitive effort, which might occur at different times. We therefore seek to address listening effort as a concept that changes over time rather than a static concept.

The focus on temporal aspects of speech perception and effort can be understood in parallel to the concerns raised by Farris-Trimble, McMurray, Cigrand, and Tomblin (2014) who noted that intelligibility scores indicate the final product of a process that includes real-time dynamics how a listener achieves the perception. Farris Trimble et al. examined various aspects of lexical access, and the current study examines the use of semantic context. In this current study, we echo their hypothesis that even when they recognize words correctly, people with cochlear implants (CIs) are likely to arrive at an end state of perception through different paths—likely in the auditory domain and potentially in the linguistic domain as well.

This study focuses on how listening effort is affected by semantic context, especially when auditory quality is degraded, and when there are potentially distracting events that might interfere with cognitive processing. Of particular interest is the time course of the benefit of using context to reduce effort while listening to a

sentence. The reason for this focus of study is the suspected dependence on context by people with hearing impairment and the suspicion that the processing of context might be slow and prolonged when the auditory signal is degraded. Multiple studies have now suggested that in individuals with CIs, linguistic processing can be delayed, perhaps occurring primarily after an utterance is over. For example, Winn (2016) found that for CI listeners, context-related reduction of pupil dilation (interpreted as a benefit of context) was late enough that, in a normal conversation, the next utterance would have already begun. Eye-tracking measures by Farris-Trimble et al. (2014) and McMurray, Farris-Trimble, and Rigler (2017) specifically show delays in commitment to lexical judgments by CI listeners and sustained attention to initial perceptions that are revised relatively late in time. Although NH listeners sometimes require extended acoustic input before recognizing short words (Grosjean, 1985), the psycholinguistics literature is replete with data suggesting rapid processing. The aforementioned CI studies suggest that language processing by this clinical population might be worth inspecting more closely to see if it deviates from the “normal” pattern in a meaningful way. If the processing of words and context by CI listeners persists far past the end of an utterance, the benefit of context might be at risk for disruption from later-occurring sounds.

A phenomenon relevant to the current study is perceptual restoration, an auditory illusion where an utterance is heard “correctly” by the listener even if it is replaced by noise or otherwise distorted (Gibson & Thomas, 1999; Samuel, 1981; Warren, 1970). The timing of this restoration has been studied in some depth; Connine, Blasko, and Hall (1991) showed that contextual information is used to disambiguate phonetic confusions when the context occurs within six syllables. Otherwise, there is delayed lexical decision and reduced benefit from surrounding context. Warren and Sherman (1974) showed that later-occurring semantic context can play an influential role in restoring missing sounds even when local acoustic cues to missing words (like coarticulation) are neutralized. It would be ideal for a listener to use context predictively, to aid in the recognition of upcoming words (Altmann & Kamide, 1999; 2007; Tavano & Scharinger, 2015), but such a process might not be recoverable from behavioral measures (Başkent et al., 2016), since a word could have been perceived correctly or restored with the aid of contextual cues, but result in a correct score either way (Samuel, 1981).

Current psycholinguistic theories of sentence processing have departed from the distinctly different notions of predictions versus retroactive restoration, instead favoring a framework where *potential* parsings of a sentence are held in partial activation until later information constrains the listener to accept a single interpretation

(Trueswell & Tanenhaus, 1994). In this view, the late use of context would not necessarily be retroactive but rather would become more solidified from the incorporation of constraints from a later piece of information, or a delayed output of language processing. Even with this modern view in mind, there is reason to suspect that in the case of hearing impairment, retroactive corrective processing would be invoked more often than in prevailing models of normal sentence processing. Specifically, we expect that hearing impairment would not only cause greater activation of potentially correct lexical competitors (cf. McMurray et al., 2017) but also activation of incorrect lexical items that are not phonological or semantic neighbors in the traditional sense. Such activation could be detrimental and necessitate corrective “garden-path”-style processing (rather than simply downstream pruning) if there is strong attachment (or “digging in” cf. Tabor & Hutchins, 2004) to the incorrect perception. Farris-Trimble et al. (2014) discuss these considerations in a study that showed increased lexical competitor activation in a group of CI listeners.

The time course of using context is under investigation here because the speed of utterances in a conversation place high demands on a listener to comprehend an utterance very quickly. Heldner and Edlund (2010) have shown that consecutive utterances in a conversation are rarely separated by silent gaps of time and often actually *overlap* in time—unlike the stimuli that we rely upon clinically and experimentally to determine accuracy for recognizing words and sentences. Heldner and Edlund’s results imply that quick comprehension and prediction are not only a luxury, but also a necessity, if a listener wants to engage in conversations at normal speeds.

A motivating principle of this study is how scientists and clinicians typically hold the view that people with hearing impairment rely more heavily on (and therefore benefit more from) context. There are a number of reasons to investigate whether this common wisdom actually generalizes to real-life scenarios. Janse and Jesse (2014) noted that the apparently greater use of context by older listeners (not necessarily with specific hearing impairment but perhaps auditorily challenged nonetheless) could result from the use of nonspeeded listening tasks, where results did not indicate “the time or effort it took to obtain the result” (p. 1844). Aydelott, Leech, and Crinion (2010) showed that the context benefit obtained by older listeners was diminished in challenging listening conditions, to a larger extent than for younger listeners. Rogers, Jacoby, and Sommers (2012) suggested that older adults’ greater reliance on context may reflect a *postperceptual bias* to respond consistently with the context, rather than their greater skill in using context during word recognition; this “false hearing” concept is akin to a similar concept articulated earlier by Samuel (1981). With these studies in mind, it is reasonable to

question whether the often-cited benefit of context (Bilger, Nuetzel, Rabinowitz, & Rzeczowski, 1984; Dubno, Ahlstrom, & Horwitz, 2000; Gordon-Salant & Fitzgibbons, 1997; Kidd & Humes, 2012; Patro & Mendel, 2016; Pichora-Fuller, Schneider, & Daneman, 1995) for people with hearing impairment might be limited to ideal laboratory settings where there is significant silent processing time after test stimuli, and therefore the use of context in real-life situations might be overestimated.

Given the observations of the aforementioned studies, we hypothesize that, in the case of a degraded auditory signal, the benefit of context will occur late enough that it will be susceptible to interference from later-occurring sounds, especially if those sounds are subject to continued auditory attention and processing as described by Backer et al. (2015). Previous studies have shown that extrinsic stimuli can interfere with sentence processing, particularly if the auditory quality of that sentence is degraded. Hunter and Pisoni (2017) found that when listeners were instructed to remember a series of digits while perceiving and repeating a sentence, intelligibility was reduced, particularly for high-context sentences where postperceptual linguistic processing could have improved the scores. Hunter and Pisoni also found that the digits—the actual source of the extrinsic load—were recalled more accurately when the target sentences were less harshly degraded, and when the intervening sentence was high context. Outside the realm of the clinical auditory sciences, there is currently lively debate over whether later occurring or extrinsic linguistic input can eradicate processing of previously heard utterances (cf. Christiansen & Chater, 2016 for a “now-or-never” framework of processing). Ongoing processing might be viewed as a “bottleneck” to incoming information, but it is also evident that processing of an utterance should also incorporate integrating it with information that comes much earlier or later in the same conversation (Bicknell, Jaeger, & Tanenhaus, 2016). So even for listeners with typical hearing in ideal conditions, the question is not whether an utterance is processed and immediately discarded, but rather whether it can be processed efficiently enough that any need for corrective action can occur under constraints of conversational speech rate. In addition, it is possible that later-occurring sounds might be not benefit comprehension of previous utterances, because they might be from nonattended talkers or simply be environmental noise. In these cases, it would be essential to avoid intrusive distraction on ongoing sentence processing. The current study will add to our understanding of this situation by focusing explicitly on the *timing* of using context to reduce listening effort, and whether context can still provide release from effort under the auditory and timing constraints described earlier.

The themes described earlier tie together into three main ideas. First, listening effort has more than just a magnitude; it has an element of *timing* that must be appreciated in order to reveal dynamics of cognitive processing that are not apparent in performance accuracy scores. Second, effort relating to corrective processing of a sentence might conflict with reception of ongoing auditory events, which might be particularly relevant for people with hearing impairment who would need to engage in corrective processing more often. Finally, we suspect that prolonged auditory processing is common in people with hearing impairment, even in the case of using context. To address these issues, this study demanded a time-series measure of listening effort that was compatible with electronic hearing devices. In the following section, we describe why pupillometry was an appropriate tool to address the issues described earlier.

### Why Pupillometry Is an Appropriate Metric for This Study

Pupillometry (measurement of changes in pupil dilation) is an ideal tool to examine the temporal dynamics of effort because it is a relatively fast physiological response that can be measured as a time series rather than an individual event. Pupillometry has been used extensively as a reliable objective index of cognitive load across a wide variety of tasks (Beatty, 1982), including speech perception (Zekveld, Kramer, & Festen, 2010) since the 1960s. When extrinsic influences (such as ambient light in the testing environment) on the pupillary response are controlled, a change in pupil diameter is a biomarker of autonomic nervous system activity, and is phase locked with general cortical activity (Reimer et al., 2016), and other signs of effort such as the cingulo-opercular system (Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014) and locus coeruleus activity (Aston-Jones & Cohen, 2005). Both the sympathetic and parasympathetic branches of the autonomic nervous system contribute to the measured response (Steinhauer, Siegle, Condray, & Pless, 2004). Beatty's (1982) review suggests that pupil responses reflect domain-general cognitive function that is not specific to just speech coding. Most importantly, pupillometry is a noninvasive technique that is unaffected by electronic devices like CIs.

Pupillometry has some limitations, particularly in the delivery of experiments, which need to be controlled in particular ways that might not apply to basic behavioral measures (see paper by Winn, Wendt, Koelewijn, & Kuchinsky, 2018, this issue). In addition to the logistical constraints on experiments, there are considerations concerning the nonlinearity of the response. Consistent with pupil dilation as a marker of effort, the dilation response does not scale linearly with the *difficulty* of a task, but rather with the amount of engagement invested by the

participant, which is a product of both difficulty and the participant's motivation and willingness to perform the task. In especially challenging situations, the pupil might show a reduced response rather than more dilation, because the participant might have disengaged. In addition, pupillary responses are known to be weaker in older participants (Winn, Whitaker, Elliott, & Phillips, 1994), requiring some careful decisions about how to handle data collected across the lifespan.

Results by Kahneman and Beatty (1966), Kahneman, Onuska, and Wolman (1968), and Hyönä, Tommola, and Alaja (1995) suggest that pupil dilation does not simply reflect auditory encoding but rather the intentional plan of the listener to process and act upon the stimulus. In these studies, some kind of mental manipulation of the stimulus was required, such as adding 1 to each number heard, dividing long stimuli into smaller chunks, or translating speech into a different language.

Challenging tasks sometimes demand not only higher peak effort but also *prolonged* effort. Prolonged effects have been present in the pupillometry literature for more than 50 years but have been given very little direct attention. Kahneman and Beatty (1966) tested mental math problems of varying difficulty, finding that the pupillary response rose to peak at roughly the same rate, the response constricted more quickly after easier problems, and remained elevated for difficult problems. Ahern and Beatty (1979) later showed that the dilation response decayed more rapidly for high-aptitude students, especially for low-difficulty problems. In other words, high-performing students spent less *time* devoting processing load to math problems and were able to capitalize on the ease of low-difficulty problems by resolving cognitive activity more quickly. A higher and sustained pupillary response is elicited for words that occur less frequently in speech (Kuchinke, Võ, Hofmann, & Jacobs, 2007), as well as for words presented along with lists of phonological competitors (Kuchinsky et al., 2013). Zekveld and Kramer (2014) found that pupil dilation in response to noise-masked sentences more quickly returned back to baseline levels when the sentences were intelligible. Conversely, for sentences of lower intelligibility, the pupil response remained elevated for a longer period of time past the end of the sentence. Similar patterns were reported by Winn, Edwards, and Litovsky (2015) who showed that elevated pupil dilation for spectrally degraded speech persisted at larger dilation for incorrect trials but shrank down for correct trials. Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld, and Kramer (2015) showed large elevated prolonged pupil dilation in cases where listeners were instructed to report two simultaneously presented sentences. Winn (2016) showed sustained dilation in response to sentences that were devoid of useful semantic context, compared with predictable high-context sentences.



Taken as a whole, these studies mentioned earlier strongly suggest that there is meaningful data to be explored in the pupillary response that continues after a sentence is heard. Piquado, Isaacowitz, and Wingfield (2010) emphasized the role of the “retention interval” (between stimulus and response) in pupillometry experiments, and the raw time-series data from many studies suggest that timing and duration of elevated listening effort should be examined—especially if we are concerned about readiness for subsequent auditory stimuli. Ahern and Beatty (1979) tested students performing mental multiplication problems, and found that students with higher aptitude showed quicker return to baseline during the retention interval, while lower aptitude students showed prolonged elevated dilation, even though both groups showed the same peak dilation and latency to peak. Motivated by this prior work, the retention interval will be a major focus in the current study, though we examine it in the focus of pupillometry data rather than the broad-focus term retention interval used by Nees (2016).

The design of this study follows that used by Winn (2016) which featured a sentence-recognition task involving high- and low-context sentences, where sentence-final words were either predictable or unpredictable, respectively. For listeners with NH, low-context sentences showed a steeper (larger) growth of pupil dilation, presumably because there was no way to know the final word, and the listeners needed sustained vigilance throughout the whole stimulus. Furthermore, low-context sentences also elicited shallower decay of the peak response (i.e., sustained elevated dilation during the retention interval) after stimulus offset, particularly when the auditory signals were degraded. A crucial measurement in that study was the timing and degree of reduction of pupillary responses for high-context sentences compared with responses for low-context stimuli. This pattern was thus interpreted to be evidence of rapid predictive mechanisms that reduced the need for vigilance, consistent with the argument by Janse and Jesse (2014) described earlier. The context-related difference in pupil size—which we regard as a beneficial “release from effort”—emerged for normal-hearing listeners *during* the sentence exposure. Conversely, listeners exposed to spectrally degraded speech, and CI listeners showed context-related reduction of pupil dilation primarily *after* the offset of the sentence, suggesting that use of context happened later, possibly reducing the opportunity for prediction of upcoming speech.

In this study, we follow up on the results observed by Winn (2016), by measuring differences in pupil dilation for high- and low-context sentences in conditions where sentences are followed by a variety of poststimulus auditory signals of varying complexity. The signals ranged from unintelligible babble to sounds with linguistic

meaning and were presented during the retention interval—the time where we suspect misperceptions would be repaired. We are essentially exploring how cognitively demanding an interfering sound can be before it disrupts the use of context to process the previous sentence. It is important to note that these postsentence disruptors impose no energetic masking on the sentences themselves; it is not “speech in noise” but rather “speech before noise.”

## Methods

### Participants

Data were collected in 40 young adults (age range: 20–40 years; average: 26) with NH; 3 were excluded from data analysis because of poor camera tracking. There were 10 adults with CIs who were older (age range: 44–87 years; average: 62). All participants were native speakers of North American English. All CI users except one acquired deafness after acquiring spoken language. CI listeners had an average of 12.5 years (median 10 years) of CI use. One CI listener chose to participate while also using a hearing aid in the contralateral ear. All but two CI listeners were bilaterally implanted; all were tested using their everyday listening settings, with both devices if applicable. No participant reported language-learning difficulties or any other cognitive problems.

### Stimuli

A 184-item subset of the 400 revised speech perception in noise (Bilger et al., 1984) sentences was used. Stimulus selection was based on the authors’ judgment of the best examples of high-context sentences, excluding those that used terminology that is now uncommon, avoiding those with emotional or evocative content (which could contaminate the pupillary response), and avoiding those whose high-context status was questionable. A random set of low-context sentences were used; the criteria for this set was less stringent. Each condition contained two lists of 23 sentences each. The ordering of high- and low-context items was pseudo-randomized; there were never more than three of the same stimulus type in a row. There was a total of 46 stimuli played to each listener, which were balanced for the number of high- and low-context items; CI listeners who participated in both condition pairs (silence & noise or digits ignored & repeated) heard twice as many.

### Conditions

There were four different conditions, corresponding to what filled the 2-s retention interval (the interim time between stimulus offset and verbal response prompt). In one condition, there were 2 s of silence in the retention

interval; this condition replicates prior experiments. In the “babble noise” condition, the interval was filled by the 8-talker babble. The third condition featured a pseudo-random three-digit sequence in the interval, which was not repeated by the listener (“digits-ignored”). The “digits-repeated” condition also contained a three-digit sequence, which was repeated along with the sentence. For all conditions, the entire target sentence—not just the final target word—was repeated.

Each poststimulus event was exactly 2 s long. The 2-s segment of babble was randomly extracted from a 1-min file, separately for each trial. All digits were modified using PSOLA in Praat to be exactly two third of a second long, for a combined 2 s for the series of three digits. The digit series was drawn randomly for each trial and never including three consecutive numbers (ascending or descending), or commonly heard sequences such as 911 or the local area code of 206. The digits and the sentences were spoken by a different talker, but after noise vocoding (described next), this was not noticeable. CI participants also did not report noticing any difference in talkers.

The postsentence events were designed to test the speculation that arose from the previous study, that the use of context occurred at least partially after the sentence. The postsentence materials introduce different amounts of interference with context processing (or with auditory cognition generally) by demanding different amounts of auditory monitoring. The silence would engage nothing; babble would engage auditory processing without engaging language processing (because it was entirely unintelligible); the ignored digits would engage language processing but not intentional attention; and the repeated digits would engage language processing and active auditory attention.

Each NH participant completed two conditions: either the silent or babble noise or the digits ignored or digits repeated. Nine of 10 CI participants completed both pairs of conditions but always on different days. Each testing block contained 23 sentences, which were a mixture of high- and low-context stimuli. There were a total of four blocks per test session, resulting in 23 sentences for each of the high- and low-context stimuli in each condition.

**Speech rate.** All sentence stimuli were slowed down by 20% using pitch-synchronous overlap-add method in the Praat software (Boersma & Weenink, 2016). This step was done to improve intelligibility for the CI listeners who sometimes report that the speech rate of the original SPIN stimuli is a bit fast. After the lengthening of stimuli, the sound quality remained natural and was not noticeably peculiar to the participants with NH.

**Noise vocoding.** For listeners with NH, we intended to provide a highly intelligible stimulus set that demanded moderate amount of difficulty, in order to engage the listeners

enough to reliably elicit pupil dilation. All stimuli (including sentences, postsentence babble, or digits) were processed with a 12-channel noise vocoder. Eight-channel vocoding usually results in intelligibility performance similar to better-performing postlingually deaf CI user performance, so 12-channel vocoding was chosen to be moderately challenging but not quite as hard as CI listening. The frequency spectrum of each stimulus was divided into 12 bands between 100 and 8000 Hz, estimated to occupy equal cochlear space (according to the function described by Greenwood, 1990). The amplitude envelope of each band was extracted and used to modulate white noise that was then filtered into the same spectral band. All 12 of the modulated noise channels were summed to produce a noise-vocoded signal that maintained the general temporal properties of the original sentence while moderately degrading the spectral resolution and temporal fine structure.

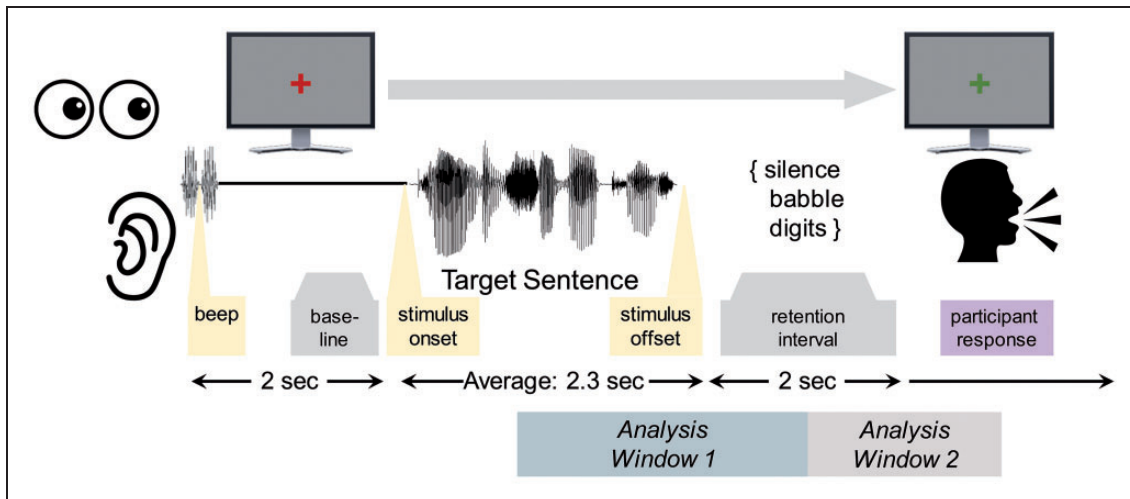
### Procedure

Participants sat in a dimly-lit double-walled soundbooth with their head stabilized on a chinrest. They were instructed to fix their gaze on a small red cross displayed on a computer monitor with a gray background. Stimuli were presented through a single loudspeaker directly in front, as their pupil size was tracked by an SR Research Eyelink eye tracker about 50 cm away from the face. Following eye-tracker calibration, there was a set of six practice stimuli played before the test phase of the experiment to familiarize the listeners with the pace of stimulus presentation.

**Trial events.** Structure of each trial is illustrated in Figure 1. Each trial began with an alerting beep, followed by 2 s of silence. One second of silence before the stimulus onset was used to establish baseline pupil diameter. The stimulus was then played, and their verbal response (repetition of the entire sentence) was elicited 2 s later via the fixation cross changing from red to green. Following the response prompt, pupil size tracking continued for another 6 s. After the end of the trial, 4 more seconds elapsed before the experimenter could manually initiate the next trial. This time was included to allow sufficient time for the pupil to return to a stable baseline level. Each trial was advanced manually by an experimenter to ensure that the next trial was presented only after allowing sufficient time for short-term changes in pupil size to subside.

### Scoring

Monitoring of eye gaze and pupil size was done live by an experimenter, who also scored accuracy of the verbal responses. Target (sentence final) words were tracked for



**Figure 1.** Sequence of events in a trial. The listener views a gray screen and fixated on a red cross. Each trial begins with an alert beep, and then 2 s of silence. Baseline pupil size was estimated in the 1 s preceding stimulus onset. After the sentence, 2 s are filled by silence, babble noise, or three random digits. Two seconds after the offset of the sentence, the cross on the screen turns green to elicit the verbal response.

accuracy. As comparison of pretarget words in the high- and low-context utterances would not yield a simple and clean keyword system, we tracked performance for words leading up to the final word as a separate single chunk, as done by Winn (2016). We also scored accuracy for the digit series in conditions where it was repeated along with the sentence. Pupil size was calculated automatically with the eye tracker.

### Data Processing

Pupil size data were marked with event tags denoting the onset and offset of stimuli. The data were cleaned of artifacts using a de-blinking procedure where stretched or missing data were expanded (to remove disturbances and mistracking at the edges of blinks) backwards by 40 ms and forwards by 80 ms and then linearly interpolated across the stretches of missing data. The data were then low-pass filtered at 10 Hz. This procedure is illustrated by Winn et al. (2018, this issue). Baseline pupil size was calculated on a per-trial basis by taking the mean pupil size during the 1 s prior to stimulus onset. All subsequent data points in the trial were scaled as proportional change relative to that baseline pupil size. Each trial was inspected to identify potential contaminations via prolonged missing data (long blinks or camera mistracking) or contamination of the baseline estimation period (as is known to happen in all pupillometry studies). We also selectively deleted brief mistracked or contaminated data (i.e., rapid pupil size disturbances outside the range of those from an evoked cognitive response, or blink artifacts not detected

at first pass) within trials, consistent with guidelines reported in previous studies, and described by Winn et al. (2018, this issue). In the case of trial contamination, 13% of trials were dropped. Within remaining trials, there was an average of 4% data lost for both NH and CI listeners. Cleaned data were aligned to stimulus offset and aggregated per listener by condition and by sentence context type. Individual aggregated data were then aggregated across the group by condition and sentence type.

### Analysis

Pupil time-series data were quantified in two ways, first to estimate proportional change in pupil size relative to baseline (with analysis done separately for each listener group to avoid confounds of pupil dynamic range), and second to estimate the relative reduction of pupil size for high-context sentences relative to peak pupil dilation obtained for low-context sentences for each listener. The second analysis was done with individual scaling of dynamic range and thus included a comparison between listener groups.

Pupil dilation was modeled using growth curve analysis (GCA), using the technique described by Mirman (2014) and used in previous pupillometry studies (Kuchinsky et al., 2013; Winn, 2016; Winn et al., 2015). GCA uses orthogonal polynomials to quantify the overall level, slope, and inflection of the time-series data. The polynomials are mapped to the range of time used for the analysis, so that there is “linear time” that is used to estimate slope, “quadratic time” used to estimate

curvature, and so on. The polynomials were equal range and scaled in time so that they were centered in the time window and were orthogonal; increases in linear growth are statistically unrelated to increases in quadratic growth (because the quadratic curve grows symmetrically, while the linear curve grows only in the positive direction). A detailed visual illustration of polynomials for GCA can be found in the supplemental materials in the article by Winn et al. (2015). We broke down the data into two analysis windows corresponding to the one-listening and two-wait or response portions of the trials, similar to the analysis by Winn (2016) for similar stimuli. The listening portion began at  $-1$  s, just after the time of sentence onset, and continued to  $+1$  s after the sentence offset, keeping in mind the approximate 1 s rise time of the pupillary response. The second window began at  $+1$  and continued to  $+2.75$  s, relative to sentence offset. In light of the aforementioned  $\sim 1$  s delay, this time window corresponds to the 2-s retention interval in preparation for the verbal response. The exact time of 2.75 s was chosen because of the distinct trough in pupil dilation at that time, corresponding to the visual change detection for the verbal response prompt. Planned fits to a window extending to 3 s produced similar results, but larger residuals at the very end of the analysis window that would only be captured by higher order polynomials. Pupil dilations later than this second window were not analyzed, as they reflect the dilation activity associated with the verbal response. Although there is likely an interesting potential question to be asked about auditory effects on verbal planning, no such effects were built into the current study.

### Quantifying Context-Based Reduction in Pupil Dilation

The benefit of context was quantified by calculating each listener's reduction in pupil dilation for high-context sentences relative to low-context sentences, divided by their peak pupil dilation for the low-context sentences, during the 3-s window following each stimulus (i.e., timepoints between 0 and 3 s). For brevity, this will be referred to as *effort release*. This analysis procedure mirrors the one done by Winn (2016) for the same comparison of contexts. This particular time window was chosen to include the typical time of the peak pupil response roughly 1 s after stimulus offset but also extended to 3 s to accommodate the later peak responses in the "digits repeated" condition. The sharp change in pupil dilation that was excluded from previous analysis did not affect the current analysis because each curve was assumed to be affected equally. An important element of this analysis is that this reduction is scaled to the dynamic range of pupil dilation observed in each listener, in each condition. If one listener shows a large amount of pupil dilation overall, a larger amount of absolute reduction is

needed to achieve the same amount of proportional "release," compared with a different listener with a smaller amount of overall dilation (see Winn et al., 2018, this issue). This measurement therefore scales with the reactivity of an individual's pupils, toward the goal of being able to normalize across listeners who have different ranges of overall dilation. Hence, we could ask questions like "given the range of dilation for sentences overall, what is the proportional reduction of that range that is obtained when context is available?" This procedure used previously by Winn (2016) was chosen to avoid simple confounds of age-related differences in dynamic range and overall stimulus-related pupil reactivity between groups. For only this measurement, statistical analysis was conducted not only across conditions but also across listener groups.

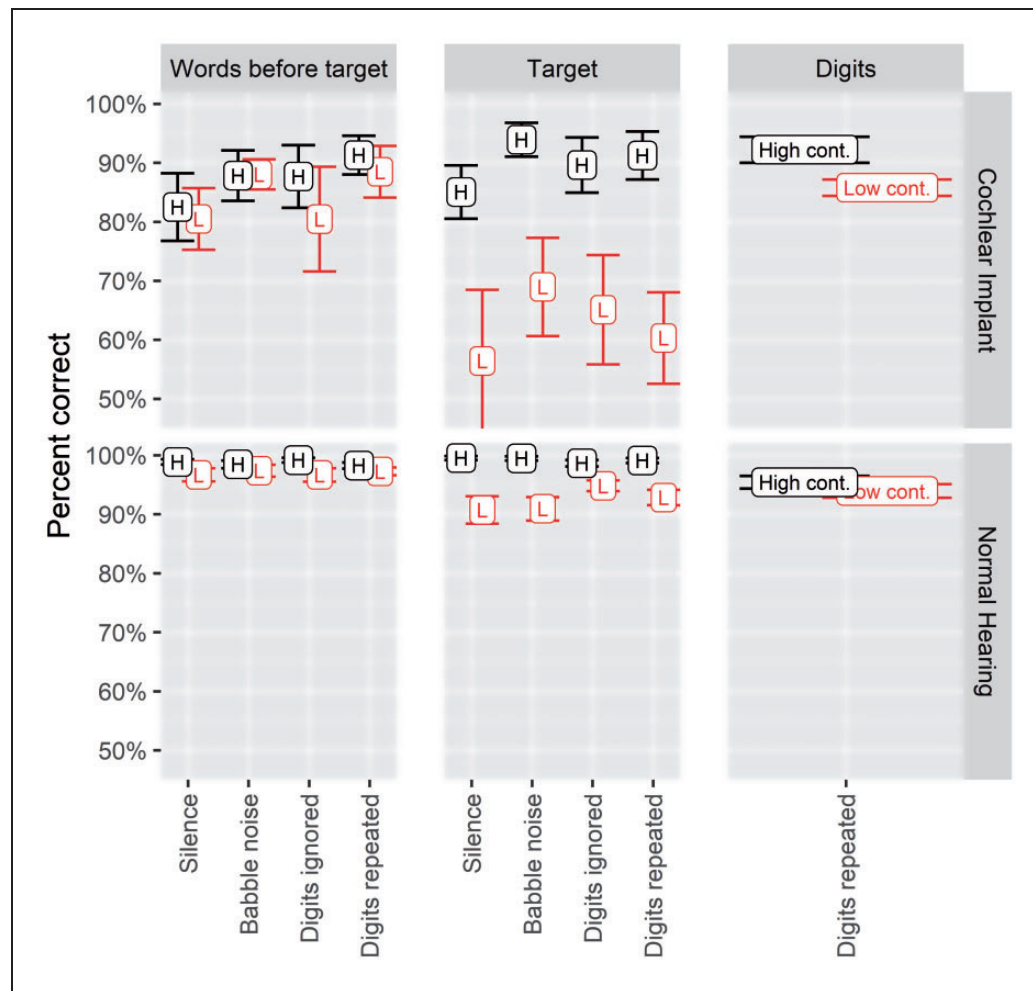
## Results

### Intelligibility

Intelligibility scores are displayed in Figure 2, broken into three categories: any words leading up to the final ("target") word in the sentence, the sentence-final word, or the digits, if in a condition where they were to be repeated. As a reminder, these sentences for NH listeners were noise vocoded so that intelligibility was high, but so that the signal still was slightly challenging. For these listeners, scores for the "context" part of the sentence (every leading to the final word) were unremarkably around 100%, and no statistical analyses were performed. Target words were reported with significantly lower accuracy in low-context sentences for the silent condition ( $\beta = -0.085$ ,  $SE = 0.0172$ ,  $p < .001$ ), and this effect was not significantly different in the "Noise" or "digits repeated" condition. In the "digits ignored" condition, the reduction of intelligibility based on context was less than that in the silent condition ( $\beta = 0.053$ ,  $SE = 0.025$ ,  $p = .052$ ) and only 3% in absolute terms.

Intelligibility scores for CI listeners were lower overall compared with scores for NH listeners, for all three categories of words. The effect of context was not observed for any words leading up to the target word, suggesting that high-context and low-context sentences were not inherently different in their word-level intelligibility. However, sentence-final target words were significantly more intelligible when preceded by relevant semantic context (i.e., in "high-context" sentences). There were two interesting and unexpected patterns in the CI group intelligibility data. First, there was no significant effect of stimulus condition (i.e., what followed the sentence) on the effect of context on the target word accuracy, despite some reliable differences in pupil dilation (seen later in Figures 3, 4, and 5). Second, the digits were reported back more accurately when preceded by





**Figure 2.** Intelligibility scores for stimuli in various conditions. “Words before target” counted as incorrect if any content words, that is, not {a/and/the/or} were repeated back incorrectly. Error bars represent  $\pm 1$  standard error of the mean. Points circled in red correspond to low-context stimuli; and black points are high-context stimuli.

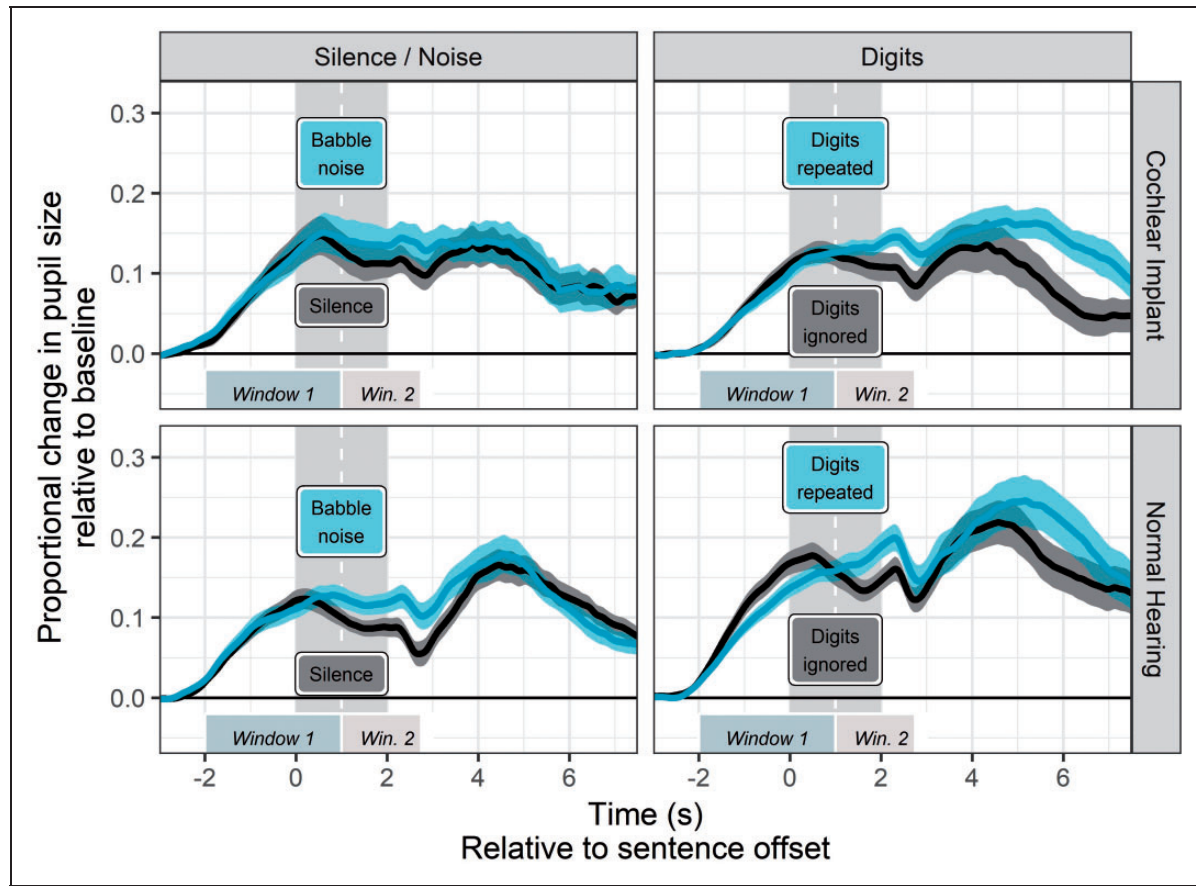
high-context sentences, at a statistical level that approaches the conventional criterion ( $p < .06$ ), suggesting that without the benefit of context, perception of subsequent speech—however simple in nature—can be disrupted.

### Pupil Dilation

The first analysis averaged over the effect of semantic context in order to focus on the effect of postsentence event. Listeners with NH showed systematically larger pupil dilation with each increase in difficulty of the postsentence portion of the stimulus. As a reminder, the target sentence stimuli in each condition were equivalent in difficulty, with the only change being what came after the sentence. In addition, listeners were always aware of which postsentence stimulus was about to be played, since the conditions were blocked. We therefore interpret

differences in pupil size to be attributable to the anticipation or processing of the postsentence stimulus.

For the purpose of clarity, Figure 3 averages over different types of context to show overall effect of condition and splits the data into pairs of conditions. This has the consequence of displaying some basic useful comparisons. On the left, the silence or babble conditions address the question of *nothing* versus *anything* after the sentence; on the right, the digits ignored or repeated address the question of *nonattended* versus *attended* stimuli. The pupil dilation shown in the figure is change in pupil size relative to baseline pupil size measured 1 s before sentence onset. Within each pair, the more difficult conditions showed greater pupil dilation (babble noise compared with silence, digits repeated compared with digits ignored). For the CI listeners, the pattern within condition pairs was reliable but was not monotonically ordered across condition pairs.



**Figure 3.** Proportional change in pupil size relative to baseline (relative to the 1 s prior to sentence onset) in response to sentences followed by various sounds or silence. Silence and babble noise are paired in left panels, and sentences followed by digits are shown on the right. In each panel, the “easy” condition (silence or digits ignored) is in black, with the “difficult” condition (noise or digits repeated) shown in light blue. Vertical gray-shaded areas represent the time between the offset of the target sentence and the prompt for verbal response. Ribbon width represents  $\pm 1$  standard error of the mean.

Specifically, there was smaller overall pupil dilation for the sentences followed by digits. This is likely due to a number of reasons, including a smaller sample size, an incomplete match of CI participant samples for these condition pairs, the overall greater variability in intelligibility among the CI group, as well as the increased likelihood of disengagement (i.e., giving up) in the more challenging conditions. Lower overall pupil dilation in hearing-impaired listener groups has been previously observed by Koelewijn, Versfeld, and Kramer (2017) and Ohlenforst et al. (2017). For that reason, we chose to display and analyze comparisons of raw pupil dilation only for condition pairs, where within-listener comparisons are appropriate.

### Growth Curve Analysis

**Window 1**—“Listening” (−2+1 s re: sentence offset). For the period of time during sentence, up to 1 s after sentence offset, the pupil dilation response reflects the “listening” phase of the trial. Because of the relatively simple shape

of the pupil data curves in this interval, the model used a quadratic function to capture the slope and basic curvature. The prevailing statistical model took the following form for each listener group:

$$\begin{aligned} \text{pupil} \sim & \text{poly1} + \text{poly2} + \text{Condition} + \\ & \text{poly1: Condition} + \text{poly2: Condition} + \\ & (1 + \text{poly1} + \text{poly2} + \text{Condition} + \text{poly1:} \\ & \text{Condition} + \text{poly2: Condition} | \text{Listener}) \end{aligned}$$

where *pupil* is the cleaned low-pass filtered proportional change in pupil size relative to baseline, and “Noise” represents the condition, that is, the postsentence auditory event (silence, babble noise, etc.). Orthogonal linear and quadratic values for time are reflected by the *poly1* and *poly2* terms. Terms in parentheses represent random effects (parameters allowed to vary within the specified group to produce a distribution that accounts for some variance in the outcome). Each of the fixed effects (including interactions) was also modeled as random effects for each listener. Degrees of freedom were

estimating using the Satterthwaite (1946) approximation using the `lmerTest` package (Kuznetsova, Brockhoff, & Christensen, 2017) in R. More complex models (higher order polynomials) did not show any detectable advantage over the quadratic model for Window 1, nor the cubic for Window 2, to be discussed later. Supplementary materials illustrate the differences in model fit with increasing polynomial order, from intercept-only models all the way to quantic (fifth order) models.

The condition term was coded with the expected “easy” condition (either silence or digits ignored) as the default (0) and the “harder” condition (noise or digits repeated) as the +1 deviant. Since separate analyses were run for the silence or noise and digits ignored or repeated, there were only two conditions in each statistical model. Separate models were run for each listener group because the differences in age are likely to have an impact on overall pupil size and dynamic range of dilation (Winn et al., 1994).

Full model results for Window 1 are shown in Table 1, and results for Window 2 are shown in Table 2.

**Sentences followed by silence or babble noise.** For both listener groups—those with NH and those with CIs—both average pupil dilation curves were virtually identical until the end of the first time window (ending at 1 s post-sentence offset), as can be observed in Figure 3. The pupil responses at the end of this time window in NH listeners were larger for sentences followed by noise, which produced similar intercept and slope terms but a smaller absolute value of the quadratic term, meaning that responses in the noise condition did not exhibit as much curvature back down toward the baseline level.

For listeners with CIs in the silence-or-noise condition pair, patterns of pupil dilation were roughly similar to those seen in NH listeners. The only difference was that the reduction of the quadratic term in CI listeners did not reach a conventional significance threshold ( $p = .09$ ) according to the generalized linear mixed model, although its estimated direction and magnitude of change were similar to that obtained in NH listeners.

**Sentences followed by digits.** When normal-hearing listeners heard sentences followed by digits, there was a statistically detectable increase in overall pupil dilation (intercept) and slope of pupil dilation when the digits were ignored compared with when the digits were repeated. The quadratic term—representing curvature or deceleration of pupil dilation—was roughly twice as large for the digits-ignored condition as for the digits-repeated condition. This reflects the pattern on Figure 3, where the curve for the digits-ignored condition starts to curve downward before time 1 s, whereas the curve for the digits-repeated condition essentially continues to rise monotonically all the way to its peak around 2.5 s

postsentence offset. This pattern was not in the same direction as the hypothesis; the condition considered to be “easier” elicited greater pupil dilation during the listening portion of the trials.

CI listeners showed a pattern for the sentence+digits conditions that resembled that of the NH listeners, although the overall change in pupil dilation (intercept term), was not statistically different across conditions. However, just as for the NH group, the CI group showed a statistically larger quadratic term for the digits-repeated condition, consistent with the downward slope of these data at the end of Window 1 (seen in Figure 3) compared with the rather continuously rising data for the digits-ignored condition in this same time window.

**Window 2—“retention interval” (+1—+2.75 s re: sentence offset).** For the period of time starting roughly 1 s after sentence offset continuing to 3 s after sentence offset, the pupil dilation response reflects the “retention interval,” or the listener’s preparation to give a verbal response after hearing the sentence (Piquado et al., 2010). Because of the multiple inflections in the pupil data curves in this interval, we modeled the data using a cubic function rather than a quadratic function. More complex functions (quartic and quantic) did not provide any statistical advantage over the prevailing cubic model. Illustration of successively more complex models is available in the Supplemental Material. For Window 2, the model took the following form:

$$\begin{aligned} \text{pupil} \sim & \text{poly1} + \text{poly2} + \text{poly3} + \text{Condition} + \\ & \text{poly1: Condition} + \text{poly2: Condition} + \text{poly3:} \\ & \text{Condition} + (1 + \text{poly1} + \text{poly2} + \text{poly3} + \\ & \text{Condition} + \text{poly1: Condition} + \text{poly2:} \\ & \text{Condition} + \text{poly3: Condition} | \text{Listener} \end{aligned}$$

where the terms are defined the same as for Window 1, with the addition of *poly3* for the cubic value for time. Again all fixed effects were also entered as random effects for each listener.

**Sentences followed by silence or babble noise.** During the retention interval, NH listeners showed statistically larger overall pupil dilation and shallower negative slope when the retention interval was filled by babble noise, reflecting greater effort that was *sustained* throughout the retention interval. In both conditions, the slope estimate was negative, but there was a steeper negative reduction (which we interpret to be beneficial release from sustained effort) in the silent condition. There was also a reduction of the quadratic term for sentences followed by babble noise that was marginally statistically detectable.

Data obtained in CI listeners during Window 2 were different from the NH data in a number of ways. First,

**Table 1.** Results of GLMM for Window 1 (Listening Portion).

	Estimate	SE	df	t	p(> t )
<b>Normal-hearing group</b>					
<i>Silence or noise</i>					
(Intercept)	0.090	0.009	16	9.75	<.001***
Condition-noise	0.003	0.007	16	0.4	.693
poly1	0.475	0.066	16	7.15	<.001***
poly2	-0.244	0.039	16	-6.22	<.001***
poly1: Condition-noise	0.035	0.052	16	0.68	.509
poly2: Condition-noise	0.110	0.047	16	2.34	.032*
<i>Digits ignored or repeated</i>					
(Intercept)	0.127	0.012	16	10.44	<.001***
Condition-digits-repeated	-0.021	0.009	16	-2.47	.025*
poly1	0.774	0.077	16	9.99	<.001***
poly2	-0.295	0.032	16	-9.12	<.001***
poly1: Condition-digits-repeated	-0.042	0.048	16	-0.87	.399
poly2: Condition-digits-repeated	0.159	0.028	16	5.6	<.001***
<b>Cochlear implant group</b>					
<i>Silence or noise</i>					
(Intercept)	0.093	0.015	9	6.09	<.001***
Condition-noise	0.003	0.010	9	0.26	.801
poly1	0.743	0.119	9	6.26	<.001***
poly2	-0.173	0.044	9	-3.95	.003***
poly1: Condition-noise	-0.016	0.059	9	-0.27	.792
poly2: Condition-noise	0.084	0.045	9	1.89	.091
<i>Digits ignored or repeated</i>					
(Intercept)	0.078	0.009	8.07	8.35	<.001***
Condition-digits-repeated	-0.004	0.005	8.03	-0.78	.46
poly1	0.703	0.058	8.08	12.04	<.001***
poly2	-0.118	0.030	8.03	-3.9	.004***
poly1: Condition-digits-repeated	-0.028	0.058	8.01	-0.48	.643
poly2: Condition-digits-repeated	0.053	0.020	8.04	2.69	.027**

Note.  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

the slope estimate in the silence condition was not statistically different from zero (unlike the negative slope for NH group); the slope for the babble noise condition was larger than that of the silence condition but only marginally detectable statistically. The interaction of slope terms resulted in an estimate of zero slope for the babble noise condition, implying that neither condition produced a detectable overall slope. The only polynomial term that was detectable in the model was the cubic term, which was statistically negative in the silence condition with smaller magnitude in the babble noise condition, again implying less change in activity in the retention interval when the interval was filled by noise.

*Sentences followed by digits.* For NH listeners, the pupil response during the retention interval began at the same magnitude but was on a considerably different course depending on whether the digits were ignored or

repeated. Essentially, the start of Window 2 was the point where the pupil dilation data were either in the middle of a fall (for digits repeated) or a rise (digits repeated). Pupil dilation for the digits-repeated condition had statistically greater overall level (intercept term) and slope, which changed sign from negative to positive. There was also a statistically larger quadratic term for the digits-repeated condition. The cubic term was not statistically detectable in either condition.

Direct statistical comparisons of NH and CI data for overall changes in pupil dilation were not made because of the inherent group differences in the dynamic ranges of pupil dilation. However, one of the patterns that was clear on the basis of gross detectable-versus-undetectable statistical differences was that nearly all (16 of 18) of the NH listeners showed an extra late peak in the pupil response for sentences followed by ignored digits, which was not observed in the CI group. It thus appears



**Table 2.** GLMM Results for Window 2 (Retention Interval).

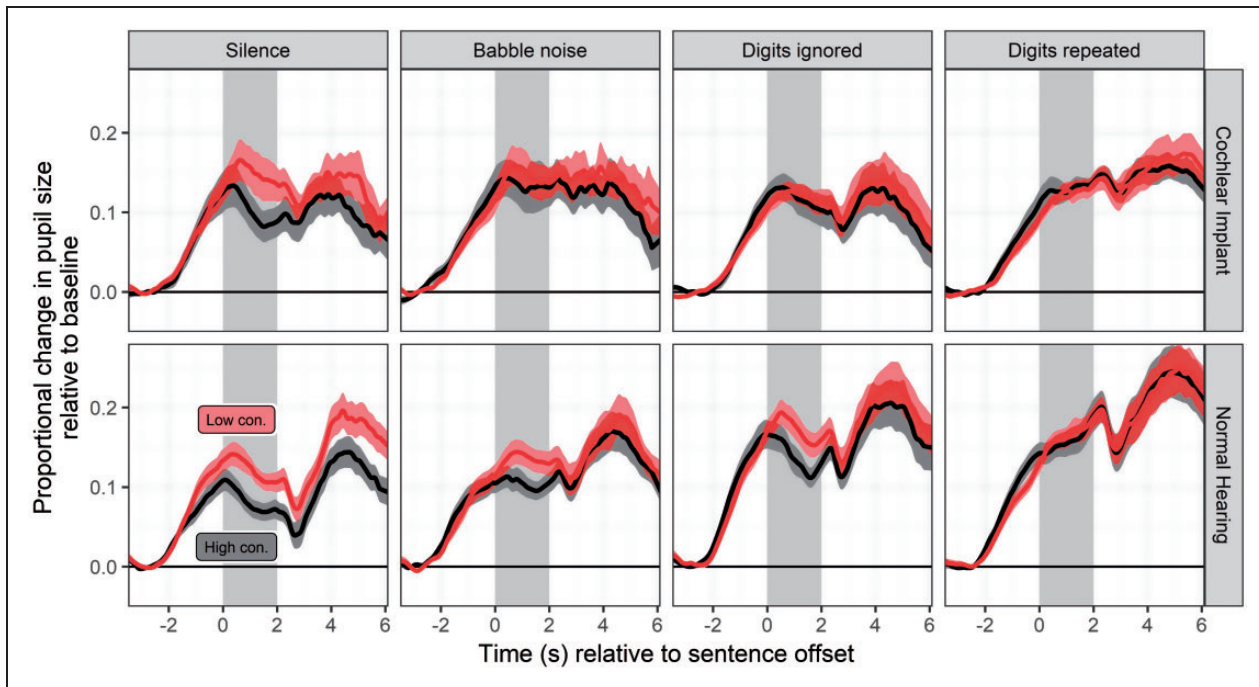
	Estimate	SE	df	t	p(> t )
Condition pair: Silence or noise					
<i>Normal-hearing group</i>					
(Intercept)	0.083	0.011	16	7.92	<.001***
Condition-noise	0.035	0.011	16	3.27	.005**
poly1	-0.128	0.063	16	-2.03	.06*
poly2	-0.060	0.022	16	-2.72	.015**
poly3	-0.060	0.012	16	-4.85	<.001***
poly1: Condition-noise	0.103	0.040	16	2.58	.02**
poly2: Condition-noise	0.045	0.024	16	1.9	.076
poly3: Condition-noise	-0.001	0.014	16	-0.07	.946
<i>Cochlear implant group</i>					
(Intercept)	0.114	0.019	9.09	5.95	<.001***
Condition-digits-repeated	0.025	0.013	8.99	1.93	.086
poly1	-0.061	0.041	9.08	-1.48	.173
poly2	0.005	0.021	8.99	0.25	.811
poly3	-0.050	0.018	9.07	-2.79	.021**
poly1: Condition-noise	0.060	0.035	9.03	1.73	.117
poly2: Condition-noise	0.012	0.017	9	0.72	.492
poly3: Condition-noise	0.025	0.010	9.03	2.39	.041*
Condition pair: Digits ignored or repeated					
<i>Normal-hearing group</i>					
(Intercept)	0.144	0.014	16	10.35	<.001***
Condition-noise	0.032	0.011	16	2.91	.01**
poly1	-0.022	0.044	16	-0.51	.62
poly2	-0.010	0.027	16	-0.35	.728
poly3	-0.120	0.019	16	-6.16	<.001***
poly1: Condition-digits-repeated	0.096	0.057	16	1.69	.111
poly2: Condition-digits-repeated	-0.112	0.028	16	-4.07	<.001***
poly3: Condition-digits-repeated	0.007	0.017	16	0.42	.678
<i>Cochlear implant group</i>					
(Intercept)	0.109	0.016	8.19	6.68	<.001***
Condition-digits-repeated	0.026	0.013	8.25	1.97	.084*
poly1	-0.115	0.055	8.04	-2.1	.069*
poly2	-0.029	0.029	8	-0.99	.349
poly3	-0.030	0.019	7.88	-1.56	.157
poly1: Condition-digits-repeated	0.169	0.041	7.67	4.1	.004***
poly2: Condition-digits-repeated	-0.004	0.033	8.19	-0.12	.91
poly3: Condition-digits-repeated	-0.009	0.014	7.92	-0.62	.554

Note.  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

that even when the auditory information is to be disregarded, the NH listeners still show some evidence of brief reduction in dilation and then a “burst” of dilation perhaps reflecting information encoding or active auditory suppression.

*The use of context to reduce listening effort.* The effect of semantic context on pupil dilation is illustrated in Figure 4 as the difference between data for low context

(red lines, usually higher) and high context (black lines, usually lower). Both sentence context types were heard in the same testing block with the same level of acoustic difficulty and the same postsentence stimulus type, so the difference between lines is attributable to the type of context. Greater difference is interpreted to reflect greater reduction of listening effort obtained through the use of context (to be discussed further). Listeners with NH who heard sentences followed by silence



**Figure 4.** Proportional change in pupil size relative to baseline, in response to high-context (easy; black lines) and low-context (difficult; red lines) sentences. Gray-shaded areas represent the time between the offset of the target sentence and the prompt for verbal response. Ribbon width represents  $\pm 1$  standard error of the mean.

replicated earlier results by Winn (2016) in that they demonstrated benefit from context even before the offset of sentences. However, that early-occurring benefit was not observed for the same listener group when sentences were followed by other auditory stimuli and not observed at all for CI listeners, also consistent with Winn's (2016) results. CI listeners showed trend in the reverse direction (faster growth of dilation for high-context sentences) in the digits-ignored and digits-repeated conditions, again with no clear explanation based on previous studies. It is possible that the small number of participants in this group resulted in a spurious effect.

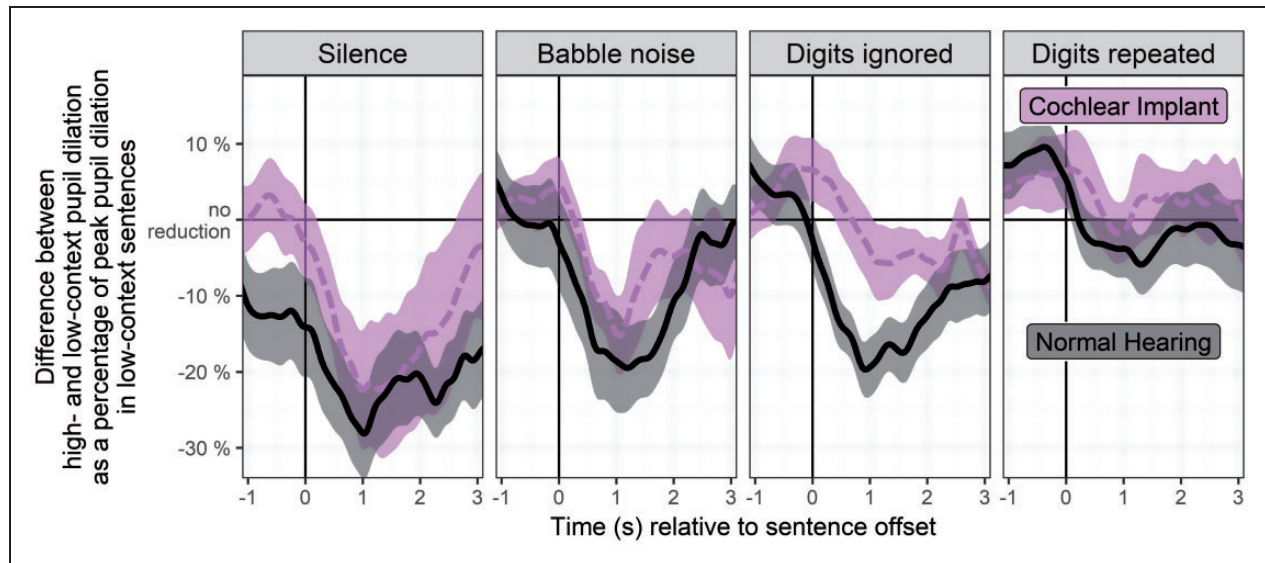
The amount of reduction in pupil size attributable to sentence context is displayed in Figure 5. As described earlier in the Analysis section, these data are normalized such that each magnitude reflected a deviation in pupil size relative to an individual's pupillary reactivity to the more-difficult (low context) stimuli; context-related change was therefore scaled to each listener's overall pupil reactivity. The range of this outcome measure is in line with that previously reported by Winn for this new sample of listeners.

Although Winn (2016) used a three-parameter sigmoid to model ongoing effort release (the difference between curves obtained for low- and high-context sentences), the data from the current study did not take sigmoidal form, and thus a different analysis approach was necessary. GCM, offering greater liberty in the

functional form of the data, was performed to model the shape of data illustrated in Figure 5. The entire data set for both NH and CI listeners was included in a unified model, and comparisons were made between groups. A cubic model was used because there were two major inflections visible for multiple data series in Figure 5. The model took the following form:

$$\begin{aligned} \text{perc\_reduction} \sim & \text{poly1} + \text{poly2} + \text{poly3} + \text{Noise} + \\ & \text{Hearing} + \text{Hearing: Condition} + \\ & \text{poly1: Condition} + \text{poly2: Condition} + \text{poly3:} \\ & \text{Condition} + \text{poly1: Hearing} + \text{poly2: Hearing} + \\ & \text{poly3: Hearing} + \text{poly1: Hearing: Condition} + \\ & \text{poly2: Hearing: Condition} + \text{poly3: Hearing:} \\ & \text{Condition} + (1 + \text{poly1} + \text{poly2} + \text{poly3} + \\ & \text{Condition} + \text{poly1: Condition} + \text{poly2:} \\ & \text{Condition} + \text{poly3: Condition} | \text{Listener}) \end{aligned}$$

The dependent variable was the percentage reduction of pupil dilation illustrated in Figure 5. The intercept term reflects the overall level of the curve; the linear term represents the slope; the quadratic term represents the curvature; and the cubic term reflects the asymmetry of the curvature or the presence of a second inflection. The main effects were noise (stimulus condition or what came after the sentence) and hearing group. Time was modeled using a third-order polynomial, with poly1, poly2, and poly3 referring to linear, quadratic, and cubic time,



**Figure 5.** Reduction of pupil dilation for high-context sentences relative to peak effort in low-context sentences. Black solid lines are data for NH listeners, and purple dashed lines are data for CI listeners. Ribbon width represents  $\pm 1$  standard error of the mean. Greater magnitude in the negative direction represents more release from effort obtained by using sentence context.

respectively. Each time parameter interacted with condition, and these interactions were the focal points of the analysis. There were subject-level random effects for each of the fixed-effect terms except for any involving hearing status (since each listener had just one hearing status).

For NH listeners in the silent condition, overall effort release (the intercept term) was greater than zero. The reduction in the intercept terms for the babble condition and the digits-ignored condition compared with silence condition were each not statistically significant. However, the intercept for the digits-repeated condition was significantly lower than that in the silence condition.

For NH listeners, the silence condition had no detectable slope, and the difference in slope was not statistically different for any of the other conditions. There was no detectable quadratic component in the silence condition, and the increase in the quadratic term in the babble condition did not reach the standard 0.05 criterion. There was no detectable change in the quadratic component in the digits-ignored or the digits-repeated condition for NH listeners compared with the same term in silence.

There was a significant cubic term in the silence condition for NH listeners, consistent with the double-deflection curve visible in Figure 5. The apparent change in the cubic component in the babble condition did not reach criterion of 0.05. The cubic components for the digits-ignored and digits-repeated conditions were not different than that for the silence condition.

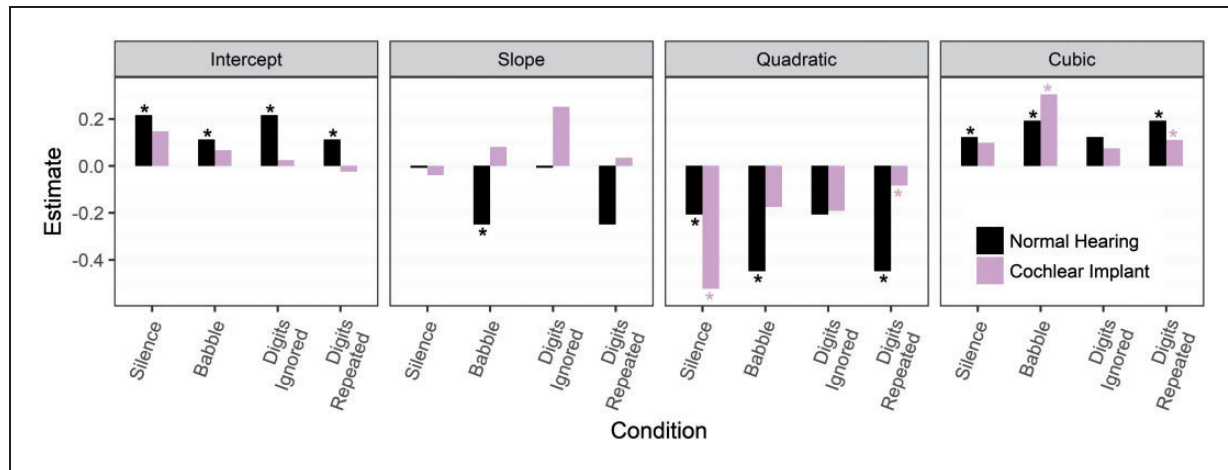
For CI listeners, the intercept terms and condition-intercept interactions were not different than those for NH listeners in each corresponding condition. The slope terms were also not different. The quadratic component

for CI listeners was different than that of the NH listeners in the digits-ignored condition, but the apparent difference in the quadratic term did not reach criterion of 0.05 for the babble condition and digits-repeated conditions.

The cubic component for CI listeners in quiet was not different than that for NH listeners. The apparent change in the cubic term for the babble noise and digits-repeated conditions did not reach criterion of 0.05, and there was no detectable difference in the cubic term for CI listeners in the digits-ignored condition. The cubic term was positive for all conditions for both listener groups, indicating greater release immediately after the sentence, which eventually declined as the participants began their verbal responses.

Illustration of the interaction-adjusted model term estimates is shown in Figure 6 to aid comparison of the numbers in Table 3. It can be seen that the overall intercept term for CI listeners is smaller than that for NH listeners and systematically declines across the four conditions ordered from least to most challenging (according to the hypothesis). There is no other pattern that emerges quite so clearly across conditions, possibly because what might appear visually as a difference in slope change might instead be best characterized mathematically as a change in quadratic or cubic term. It is possible that alternative methods of curve analysis might be more informative than the methods used here.

As for the previous analyses of overall pupil dilation, we encourage caution in drawing conclusions based on direct comparisons of the NH and CI groups in light of the differences in age and small-sample size in the CI



**Figure 6.** Model estimates for the proportional reduction of pupil dilation illustrated in Figure 5 and listed in Table 3. Bar height indicates magnitude of the model term for each condition separated by hearing group. Stars indicate that the estimate was statistically different from zero.

group. However, even without direct statistical comparison, it should be noted that the amount of context-related reduction in pupil dilation was at least statistically detectable for the NH group in all conditions but not statistically different from zero for the CI group except in the ideal silent condition.

## Discussion

In the current study, we examined how listening effort—indexed by changes in pupil dilation—is affected by the presence of auditory stimuli that occur after a sentence is over (see Figure 3). Pupil dilation also was affected by the presence or absence of semantic context in the sentences. In ideal conditions of a sentence in quiet followed by silence, context was associated with smaller pupil dilation—especially for NH listeners, and also to a smaller extent for older CI listeners (see Figures 4 and 5, left panels in each figure). The NH listeners demonstrated this benefit rapidly, even before the sentence was over, while the CI listeners did not. These results were consistent with the data published by Winn (2016) and consistent with the concept of more rapid use of context in NH listeners.

The current study also shows that the benefit of context to reduce effort is susceptible to interference by later-occurring sounds like noise or simple sequences of digits—especially if the use of context occurs primarily later in time, as for the current group of older CI listeners (Figures 5 and 6). These results reinforce the idea that the perception and successful comprehension of speech depend on cognition that continues well past the sensory perception of the signal itself—particularly in the case of a listener with HL. We further suspect that the interference with the use of context might be more

pronounced if the stimulus after the sentence were more cognitively challenging than just three digits. Although we expected that the context-related reduction in dilation for NH listeners would be robust to later-occurring sounds, it too was susceptible to later interference. Surprisingly, the reduction of context benefit as measured by *pupil dilation* did not affect the benefit of context for intelligibility in this task (Figure 2). However, there was a measurable effect of reduced accuracy in reporting the digits when the preceding sentence was low context (Figure 2, upper right panel). In a situation with more challenging material, it is possible that intelligibility effects might emerge more strongly.

Progressively more challenging events that occur after a sentence (silence, babble, ignored digits, and repeated digits) resulted in systematically higher *overall* peak pupil responses in NH listeners (Figure 3). Furthermore, the timing of the peak was distinctly different depending on if the listeners were instructed to attend to a signal after the sentence. Results for CI listeners were less clear on this pattern, possibly because of nonmonotonicity of the pupillary response (i.e., when the task is very challenging, the response will shrink rather than grow, because the listener might disengage), or possibly because of the impact of aging on pupil reactivity. The nonmonotonicity and general reduction of the pupil dilation response in listeners with hearing impairment is consistent with other work (Koelewijn et al., 2017). It is also a possibility that the relatively smaller number of CI listeners (compared with the NH group) was responsible for the less-clear pattern of results and the general lack of robust statistically identified effects.

Both the current study and the one it is based on (Winn, 2016) are limited in that there were considerable



**Table 3.** LMER Model Summary for Data Displayed in Figure 5.

	Estimate	SE	df	t	p(> t )
Normal-hearing group effects					
Intercept (silence)	0.221	0.053	22.2	4.16	.000***
Intercept: Condition-babble	-0.091	0.062	20	-1.47	.157
Intercept: Condition-digits-ignored	-0.084	0.061	34.72	-1.39	.173
Intercept: Condition-digits-repeated	-0.198	0.063	33.84	-3.16	.003**
poly1	0.079	0.146	37.23	0.54	.592
poly1: Condition-babble	-0.113	0.163	33.97	-0.69	.493
poly1: Condition-digits-ignored	-0.038	0.186	52.79	-0.21	.837
poly1: Condition-digits-repeated	-0.057	0.195	45.99	-0.29	.773
poly2	-0.166	0.130	32.56	-1.28	.209
poly2: Condition-babble	-0.272	0.162	33.57	-1.67	.103
poly2: Condition-digits-ignored	-0.158	0.153	45.94	-1.03	.308
poly2: Condition-digits-repeated	0.005	0.154	46.43	0.04	.972
poly3	0.158	0.056	15.3	2.81	.013*
poly3: Condition-babble	-0.143	0.092	14.85	-1.56	.140
poly3: Condition-digits-ignored	-0.078	0.070	32.05	-1.12	.272
poly3: Condition-digits-repeated	-0.089	0.080	28.08	-1.11	.276
Cochlear implant group effects					
Intercept (silence): CI	-0.059	0.087	22.2	-0.68	.503
Intercept: Condition-babble : CI	-0.007	0.102	20	-0.07	.949
Intercept: Condition-digits-ignored : CI	-0.040	0.098	26.57	-0.41	.686
Intercept: Condition-digits-repeated : CI	0.006	0.110	22.9	0.06	.956
poly1: CI	0.122	0.239	37.23	0.51	.613
poly1: Condition-babble : CI	-0.025	0.268	33.97	-0.09	.927
poly1: Condition-digits-ignored : CI	0.109	0.275	30.97	0.40	.695
poly1: Condition-digits-repeated : CI	-0.198	0.274	31.54	-0.72	.474
poly2: CI	-0.223	0.213	32.56	-1.05	.303
poly2: Condition-babble : CI	0.376	0.267	33.57	1.41	.169
poly2: Condition-digits-ignored : CI	0.442	0.208	36.58	2.12	.041*
poly2: Condition-digits-repeated : CI	0.359	0.191	20.76	1.88	.074
poly3: CI	-0.099	0.092	15.3	-1.07	.301
poly3: Condition-babble : CI	0.297	0.150	14.85	1.98	.067
poly3: Condition-digits-ignored : CI	0.042	0.090	22.88	0.46	.649
poly3: Condition-digits-repeated : CI	0.187	0.104	15.21	1.80	.092

Note.  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

age differences between the patient population and the NH control groups. Effects of age in working memory and speech perception have been previously documented by Zekveld, Rudner, Johnsrude, and Ronnberg (2013) and Gordon-Salant and Cole (2016). However, when controlling for age in the NH population, a change in stimulus degradation produced results that were similar to the CI results in that the benefit of context was reduced and delayed. Based on that prior observation, we are confident in drawing attention to the potential risk of delayed context benefit in the current CI test group, as the delay can be explained by auditory degradation even without any aging effects.

### The Timing of Pupil Dilation Peaks

The pattern of greater increase in pupil dilation for digits that were *ignored* compared with those that were repeated (for NH listeners) was unexpected. However, this could be reframed as simply a difference in rise time to peak dilation, which normally occurs at the end of the relevant (attended) stimulus. Such a difference in peak latency is consistent with other previous studies where listeners knew to expect shorter or longer stimuli. For example, different lengths of digit sequences tested by Klingner, Tversky, and Hanrahan (2011) produced the same slope of pupil dilation growth, but different

dilation onset and peak times entirely consistent with the length of the auditory stimulus. In addition, data presented by Borghini (2017) showed a similar change in time to peak dilation for sentences spoken at different rates of speech. If dilation at peak time represents the greatest deployment of cognitive resources, then the two different digits conditions demand different time courses of resource allocation. Sentences followed by ignored digits demand all resources deployed earlier in time, while sentences followed by repeated digits demand a more prolonged deployment of attentional resources.

By invoking the resource allocation framework of Cheng (1985), we can describe why the pupil dilation for sentences in the digits-ignored condition was higher than that for sentences in the digits-repeated condition; when digits become capacity free (because you can ignore them), then attention can be *restructured* to be devoted to other components, like the target sentence. When the digits demand capacity, less is remaining for the sentence, and so pupil dilation was accordingly smaller for the sentence when attention was spread out to include the digits. The status of a task as being capacity free or capacity demanding is dynamic and apparently under voluntary control, as the same listeners participated in both conditions but demonstrated distinctly different patterns of pupil dilation consistent with stimulus expectations.

Even when the duration of an incoming stimulus is not known explicitly, the temporal aspects of pupil dilation can be inspected to reveal specific aspects of cognitive processing. A study by Bradshaw (1968) suggested that prolonged pupil dilation responses are an indication that a participant is not finished solving a problem. Peak pupillary responses in that study were aligned to problem-solution time, indicated explicitly by the participants performing math problems. When answers were cued in a predictable manner, peak dilation aligned with the extrinsic cue. Hence, problem solving, preparation for response, and planned timing of the response all appear to drive peak dilation latency.

### *Interpreting (Lack of) Intelligibility Effects*

Even when the context-related effort release (reduction in pupil size for high-context sentences) was diminished by poststimulus auditory events, context still provided increase in intelligibility (Figure 2). Clearly then, postsentence auditory stimuli do not entirely eradicate context benefit. However, we note that the long-term comprehension and memory of spoken language depends on more than just accuracy at the time of perception. Whatever resources that are needed for encoding into memory might be compromised if they are used for competing

tasks. The *effortfulness hypothesis* (McCoy et al., 2005) and the ease of language understanding model (Rönnberg, Rudner, Foo, & Lunner, 2008; Rönnberg et al., 2013) both describe a framework of limited resources for storing speech into memory, which are further diminished if extra resources are needed for the initial sensory processing of speech. In other words, an auditory signal that is degraded should produce worse encoding and recall than a clear signal. This prediction has been validated in numerous studies McCoy et al. (2005) showed impaired word recall in older adults if they had mild HL. Several studies (Cousins, Dar, Wingfield, & Miller, 2014; Rabbitt, 1968) have shown that auditory degradation such as noise masking impairs recall of auditory word stimuli, even if the words were recognized correctly at the time of perception. Other long-term effects were identified by Gilbert, Chandrasekaran, and Smiljanic (2014), who found that clearer speech signals were remembered more accurately, in terms of correct recall and correct rejections of items not heard. Van Engen, Chandrasekaran, and Smiljanic (2012) similarly found improved recognition memory for easier-to-process meaningful and clear sentences, suggesting that the absence of pupil size reduction for high-context sentences in the current study might not have neutralized intelligibility differences in the short term but instead might have still potentially reduced memory encoding in a way that was not measured in this study. We cannot know for sure based on the current testing paradigm.

### *Context Benefit: What Are We Measuring?*

In light of the current results, it is worth considering how the usefulness of context is measured. The prevalence of context effects (i.e., the use of context to support better intelligibility scores) in many studies shows that context is beneficial when available. However, if a listener habitually uses context to retroactively “repair” perception of a degraded auditory signal, and later-occurring auditory stimuli can interfere with that repair process, the benefit of context might not survive outside the confines of an idealized laboratory or clinical environment. The quickness of comprehension should be crucial in conversational settings, where prolonged cognitive activity devoted to processing a recent sentence could jeopardize a listener’s readiness to understand the next sentence. As shown by Hsu and Novick (2016), attentional control for cognitively engaging tasks (in their case, the Stroop test) appears to adversely affect processing a later-heard sentence. Consistent with the results of Aydelott et al. (2010), we show that the benefit of context can be diminished in challenging listening conditions—perhaps just where it is needed the most. We support the notion described by Janse and Jesse (2014) that the measured benefit of context in many prior studies might have

resulted from the use of nonspeeded listening tasks. Because typical conversations rarely include silent time that a listener could use to “catch up” after making a perceptual mistake, the results of the current study might reflect an important part of what makes listening effortful when one has HL.

Even though the current study focuses on rather short-term (within 2–3 s) use of context, it is not because later-occurring context is not beneficial for speech comprehension. Later-occurring information can disambiguate utterances heard much farther back in time, as is probably a common experience among many people. However, of particular interest in this investigation is not so much the ultimate comprehension of a message, but the cognitive energy exerted during listening, and what cues are used to reduce that mental load in real time. These concerns parallel those mentioned earlier by Farris-Trimble et al. (2014). To the extent that we are interested in listening effort and how to alleviate it in people with HL, it appears that one of the sources of elevated effort might be a deficit in the ability to use effort to *rapidly* decrease effort as an utterance is heard.

It is not surprising to continually be reminded by patients and clinicians alike that conventional standardized single-utterance speech perception measures do not fully describe real-life difficulty of HL. Speech communication includes multiple talkers, spatial locations, conversation topics, and distracting noises. The listener is often also a talker and might also be engaged in other activities like thinking about upcoming conversation topics, remembering recent events, and planning their responses. All of these factors could jeopardize any deliberate dedicated processing of a single sentence, yet single utterances are the norm when evaluating speech intelligibility. Individuals with HL sometimes report that when a single sentence is misunderstood, comprehension of following sentences can be derailed. To more thoroughly understand the abilities and difficulties of people with HL, more studies could be designed that explicitly aim at describing the process of ongoing auditory and cognitive processing as it would occur in the case of multiple utterances, rather than single phoneme or single-word, or even single-sentence materials. It remains unclear just how much conventional speech recognition measures might overestimate a person's ability to recognize auditory signals versus reconstruct them using realistic linguistic constraints. We recommend that in addition to considering the outcomes of experiments of speech *in* noise, it might be worthwhile to explicitly study the perception of speech *before* noise as well, to selectively examine the role of repair processes that could indicate poor auditory performance and susceptibility to interference in everyday environments.

## Conclusions

Semantic context within a sentence reduces task-evoked pupil dilation and therefore appears to reduce listening effort in listeners with NH and in listeners with CIs. The release from effort—indexed here by reduced pupil dilation for high-context sentences—is stronger and earlier in listeners with NH. The presence of auditory stimuli after a target sentence weakens the release from effort, particularly for CI listeners, whose context benefit is primarily observed *after* the offset of a sentence. Lack of contextual cues also led to poorer perception of subsequent speech (in this case, a series of digits) in CI listeners. It remains unclear whether prolonged elevated effort is directly the result of having to repair individual words but that could be explored with methods used in previous studies of perceptual restoration (Warren, 1970) or elliptical speech (Herman & Pisoni, 2003).

Listeners with NH appear showed pupil dilation responses that grew larger and smaller at specific time points, suggesting that they might be able to allocate cognitive effort in specific points in time, evidenced by greater effort earlier for the case of a sentence followed by ignored digits, and greater effort later in the case of a sentence followed by attended digits. In CI listeners, this pattern was less clear.

Previous studies highlight the benefit of context in people with hearing impairment have presented one stimulus at a time, in an unspeeded task. It is possible—even *likely*—that listeners in those studies were able to construct their response retroactively to be well formed, rather than reporting the results of an accurate perception. Context might not be exploited quickly enough by CI listeners in consecutive sentences at conversational speed. Future work might explore that possibility by using multiple sentences or postsentence stimuli that are more complex than three digits, or by altering the speed of information transmission by reducing rate of speech.

## Author's Note

Portions of this work were presented at the Conference on Implantable Auditory Prostheses (Lake Tahoe, CA; 2017) and the Pupillometry in Hearing Science Workshop (Amsterdam; 2017).

## Acknowledgments

The authors appreciate the time given by our participants to contribute their time and ideas to this project. Moira McShane and Tiffany Mitchell assisted with data collection. We also thank various colleagues for helpful suggestions on the design and interpretation of the results, including Daniel McCloy, Franzo Law II, Chad Ruffin, and Jay Rubinstein.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research 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 research was funded by NIH-NIDCD 5R03DC014309 (to M. W.).

## Supplemental Material

Supplemental material for this article is available online.

## References

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude. *Science*, 205, 1289–1292. doi:10.1126/science.472746
- Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye-movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439. doi:10.1006/jmla.1997.2558
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. doi:10.1016/S0010-0277(99)00059-1
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518. doi:10.1016/j.jml.2006.12.004
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450. doi:10.1146/annurev.neuro.28.061604.135709
- Aydelott, J., Leech, R., & Crinion, J. (2010). Normal adult aging and the contextual influences affecting speech and meaningful sound perception. *Trends in Amplification*, 14, 218–232. doi:10.1177/1084713810393751
- Backer, K., Binns, M., & Alain, C. (2015). Neural dynamics underlying attentional orienting to auditory representations in short-term memory. *The Journal of Neuroscience*, 35, 1307–1318. doi:10.1523/JNEUROSCI.1487-14.2015
- Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., & Gaudrain, E. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging. *Trends in Hearing*, 20, 1–16. doi:10.1177/2331216516670279
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. doi:10.1037/0033-2909.91.2.276
- Bicknell, K., Jaeger, T. F., & Tanenhaus, M. K. (2016). Now or...later: Perceptual data is not immediately forgotten during language processing. *Behavioral and Brain Sciences*, 39, 23–24. doi:10.1017/S0140525X15000734
- Bilger, R., Nuetzel, J., Rabinowitz, W., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, 27, 32–48. doi:10.1121/1.2017541
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.18) [Computer program]. Retrieved <http://www.praat.org/>
- Borghini, G. (2017). *Listening effort during speech understanding in a second language*. Paper presented at the Pupillometry in Hearing Science Workshop, Amsterdam, The Netherlands.
- Bradshaw, J. L. (1968). Pupil size and problem solving. *Quarterly Journal of Experimental Psychology*, 20(2), 116–122. doi:10.1080/14640746808400139
- Cheng, P. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, 92, 414–423. doi:10.1037/0033-295X.92.3.414
- Christiansen, M., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral Brain Science*, 39, e62. doi:10.1017/S0140525X1500031X
- Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, 30(2), 234–250. doi:10.1016/0749-596X(91)90005-5
- 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(4), 622–638. doi:10.3758/s13421-013-0377-7
- Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *Journal of the Acoustical Society of America*, 107(1), 538–546. doi:10.1121/1.428322
- Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, 11(1), 31–46. doi:10.1177/1084713806298004
- Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation: Cochlear implant users and normal-hearing listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 308–327. doi:10.1037/a0034353
- Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention Perception and Psychophysics*, 72, 501–516. doi:10.3758/APP.72.2.501
- Francis, A. L., MacPherson, M. K., Chandrasekaran, B., & Alvar, A. M. (2016). Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology*, 7, 1–15. doi:10.3389/fpsyg.2016.00263
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3), 225–248. doi:10.1080/016909699386293
- Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America*, 135(1), 389–399. doi:10.1121/1.4838975
- Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear*



- and Hearing, 37(5), 593–602. doi:10.1097/AUD.0000000000000316
- Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech Language and Hearing Research, 40*, 423–431. doi:10.1044/jslhr.4002.423
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species – 29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592–2605. doi:10.1121/1.399052
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics, 38*, 299–310. doi:10.3758/BF03207159
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*(4), 555–568. doi:10.1016/j.wocn.2010.08.002
- Herman, R., & Pisoni, D. (2003). Perception of “elliptical speech” following cochlear implantation: Use of broad phonetic categories in speech perception. *Volta Review, 102*, 321–347.
- Hicks, C., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language and Hearing Research, 45*, 573–584. doi:10.1044/1092-4388(2002)046
- Hoffman, H., Dobie, R., & Losonczy, K. (2017). Declining prevalence of hearing loss in US adults aged 20 to 69 years. *The Journal of the American Medical Association Otolaryngology Head and Neck Surgery, 143*, 274–285. doi:10.1001/jamaoto.2016.3527
- Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science, 27*(4), 572–582. doi:10.1177/0956797615625223
- Hunter, C. R., & Pisoni, D. B. (2017). Extrinsic cognitive load impairs spoken word recognition in high- and low-predictability sentences. *Ear and Hearing, 39*, 378–389. doi:10.1097/AUD.0000000000000493
- Hyönä, J., Tommola, J., & Alaja, A. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48*, 598–612. doi:10.1080/14640749508401407
- Janse, E., & Jesse, A. (2014). Working memory affects older adults’ use of context in spoken-word recognition. *Quarterly Journal of Experimental Psychology, 67*(9), 1842–1862. doi:10.1080/17470218.2013.879391
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583–1585. doi:10.1126/science.154.3756.1583
- Kahneman, D., Onuska, L., & Wolman, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Quarterly Journal of Experimental Psychology, 20*(3), 309–311. doi:10.1080/14640746808400168
- Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *The Journal of the Acoustical Society of America, 131*, 1434. doi:10.1121/1.3675975
- Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology, 48*(3), 323–332. doi:10.1111/j.1469-8986.2010.01069.x
- Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research, 323*, 81–90. doi:10.1016/j.heares.2015.02.004
- Koelewijn, T., Versfeld, N. J., & Kramer, S. E. (2017). Effects of attention on the speech reception threshold and pupil response of people with impaired and normal hearing. *Hearing Research, 354*, 56–63. doi:10.1016/j.heares.2017.08.006
- Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing, 33*(2), 291–300. doi:10.1097/AUD.0b013e3182310019
- Kramer, S. E. (2008). Hearing impairment, work, and vocational enablement. *International Journal of Audiology, 47*(Suppl. 2): S124–S130. doi:10.1080/14992020802310887
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspect of auditory handicap by means of pupil dilation. *Audiology, 36*, 155–164.
- Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam checklist for hearing and work. *International Journal of Audiology, 45*, 503–512. doi:10.1080/14992020600754583
- Kuchinke, L., Võ, M. L. H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65*(2), 132–140. doi:10.1016/j.ijpsycho.2007.04.004
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., 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
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. doi:10.18637/jss.v082.i13
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. doi:10.1097/00003446-199802000-00001
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243. doi:10.1016/j.cogpsych.2009.04.001
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults’ memory for speech. *The Quarterly Journal of Experimental Psychology A, Human Experimental Psychology, 58*(1), 22–33. doi:10.1080/02724980443000151
- McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less

- incrementally. *Cognition*, *169*, 147–164. doi:10.1016/j.cognition.2017.08.013
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. New York, NY: CRC Press.
- Morata, T., Themann, C., Randolph, R., Verbsky, B., Byrne, D., & Reeves, E. (2005). Working in noise with a hearing loss: Perceptions from workers, supervisors, and hearing conservation program managers. *Ear and Hearing*, *26*(6), 529–545. doi:10.1097/01.aud.0000188148.97046.b8
- Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology*, *7*, 1–6. doi:10.3389/fpsyg.2016.01892
- 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
- Patro, C., & Mendel, L. L. (2016). Role of contextual cues on the perception of spectrally reduced interrupted speech. *Journal of the Acoustical Society of America*, *140*, 1336–1345. doi:10.1121/1.4961450
- Pichora-Fuller, K., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608. doi:10.1121/1.412282
- 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
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolia, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*, 13289. doi:10.1038/ncomms13289
- Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, *27*, 33–45. doi:10.1037/a0026231
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 1–17. doi:10.3389/fnsys.2013.00031
- Rönnerberg, 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. doi:10.1080/14992020802301167
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474. doi:10.1037/0096-3445.110.4.474
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110–114. doi:10.2307/3002019
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, *52*(1), 77–86. doi:10.1016/j.ijpsycho.2003.12.005
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 431–450. doi:10.1037/0278-7393.30.2.431
- Tavano, A., & Scharinger, M. (2015). Prediction in speech and language processing. *Cortex*, *68*, 1–7. doi:10.1016/j.cortex.2015.05.001
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 155–179). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS One*, *7*(9), e43753. doi:10.1371/journal.pone.0043753
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(917), 392–393. doi:10.1126/science.167.3917.392
- Warren, R. M., & Sherman, G. (1974). Phonemic restorations based on subsequent context. *Perception and Psychophysics*, *16*, 150–156. doi:10.3758/BF03203268
- Winn, B., Whitaker, D., Elliott, D., & Phillips, J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, *35*, 1132–1137.
- Winn, M., Wendt, D., Koelewijn, T., & Kuchinsky, S. (2018). Best practices in using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1–32. doi:10.1177/2331216518800869
- Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*, 1–17. 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), e153–e165. doi:10.1097/AUD.0000000000000145
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, *101*, 76–86. doi:10.1016/j.neuroimage.2014.06.069
- Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*, 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*(4), 480–490. doi:10.1097/AUD.0b013e3181d4f251
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnerberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *Journal of the Acoustical Society of America*, *134*, 2225–2234. doi:10.1121/1.4817926.