

Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants

Trends in Hearing
2016, Vol. 20: 1–17
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516669723
tia.sagepub.com


Matthew Winn¹

Abstract

People with hearing impairment are thought to rely heavily on context to compensate for reduced audibility. Here, we explore the resulting cost of this compensatory behavior, in terms of effort and the efficiency of ongoing predictive language processing. The listening task featured predictable or unpredictable sentences, and participants included people with cochlear implants as well as people with normal hearing who heard full-spectrum/unprocessed or vocoded speech. The crucial metric was the growth of the pupillary response and the reduction of this response for predictable versus unpredictable sentences, which would suggest reduced cognitive load resulting from predictive processing. Semantic context led to rapid reduction of listening effort for people with normal hearing; the reductions were observed well before the offset of the stimuli. Effort reduction was slightly delayed for people with cochlear implants and considerably more delayed for normal-hearing listeners exposed to spectrally degraded noise-vocoded signals; this pattern of results was maintained even when intelligibility was perfect. Results suggest that speed of sentence processing can still be disrupted, and exertion of effort can be elevated, even when intelligibility remains high. We discuss implications for experimental and clinical assessment of speech recognition, in which good performance can arise because of cognitive processes that occur after a stimulus, during a period of silence. Because silent gaps are not common in continuous flowing speech, the cognitive/linguistic restorative processes observed after sentences in such studies might not be available to listeners in everyday conversations, meaning that speech recognition in conventional tests might overestimate sentence-processing capability.

Keywords

speech perception, listening effort, predictive processing, cochlear implants, context

Date received: 25 January 2016; revised: 26 August 2016; accepted: 26 August 2016

Introduction

Cognition can be described as a continuous state of prediction (Bar, 2007; Clark, 2013; Lupyan & Clark, 2015), and people demonstrate the ability to predict spoken language in a variety of ways (Tavano & Scharinger, 2015). Language predictions arise from low-level acoustic details as well as higher level processing of language content. This study aims to explore the ability to create predictions and exploit them to reduce listening effort, and how this process is affected by the spectral quality of the auditory signal input.

Predictions in speech perception are commonly revealed through measurements of eye gaze. For example, when hearing a sentence beginning with “the boy

will eat . . .,” observers will direct their gaze more quickly to edible objects like a cake and away from inedible objects like a vase (Altmann & Kamide, 1999). This effect is sensitive to subtle details of visual objects and their compatibility with ongoing speech; when seeing both a full and partially full glass, an observer hears “the man has drunk . . .” and will proceed to look more quickly at the partially full glass, as the full glass

¹Department of Speech & Hearing Sciences, University of Washington, Seattle, WA, USA

Corresponding author:

Matthew Winn, Department of Speech & Hearing Sciences, University of Washington, Seattle, 1417 NE 42nd Street, Seattle, WA 98105, USA.
Email: mwinn2@uw.edu



could not have been drunk (Altmann & Kamide, 2007). Context also improves basic word intelligibility; when hearing the sentence “She made the bed with clean . . .,” the word “sheets” is highly predictable and intelligible, even if masked by noise (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Pichora-Fuller, Schneider, & Daneman, 1995).

In addition to drawing upon our semantic knowledge, predictive processing exploits low-level acoustic details of the signal. It is possible to predict upcoming consonant sounds according to subtle differences in earlier occurring vowel sounds (a phenomenon called coarticulation). Both adults (Gow, 2002) and toddlers (Mahr, Mcmillan, Saffran, Weismer, & Edwards, 2015) can exploit coarticulation in order to more quickly process spoken words. Speech processing is affected by a variety of dimensions arising from both the content of the message and the local acoustics, as well as indexical properties of the talker; all of these factors have been shown to interact with signal *quality*, such that individuals with hearing loss might exploit context in ways that are different from the typical hearing population. The current study explores the extent to which auditory quality affects the use of context, particularly as it can reduce listening effort.

Cochlear Implants and Signal Degradation

Cochlear implants (CIs) are neural prostheses designed to electrically stimulate the auditory nerve in order to provide hearing sensation to those who have severe-to-profound hearing loss and who elect to communicate in the oral/aural mode. In this clinical population, good speech recognition in quiet is now common (Blamey et al., 2013) despite notoriously poor signal quality in the spectral domain, which results from broad (rather than tonotopically precise) spread of electrical activity within the cochlea (Boëx, de Balthasar, Kós, & Pelizzone, 2003). Degradations in signal quality have detrimental effects on perception of speech in noise (Fu & Nogaki, 2005) and also yield atypical patterns of phonetic cue weighting in the auditory (Moberly, Lowenstein, & Nittrouer, 2016; Winn, Chatterjee, & Idsardi, 2012) and visual (Winn, Rhone, Chatterjee, & Idsardi, 2013) domains. The spectrally degraded signal generated by a CI can be expected to reduce the quality of speech information that might be used to predict upcoming speech, as words are more likely to be misperceived or perceived according to some atypical listening strategy.

The Importance of Listening Effort

Listening effort can be defined as the mental exertion required to attend to and understand an auditory

message (McGarrigle et al., 2014). This topic has garnered special attention from audiologists and experimenters concerned with the performance of people with hearing loss, who routinely demonstrate elevated listening effort (Kramer, Kapteyn, Festen, & Kuik, 1997), even if the hearing impairment is mild (McCoy et al., 2005; Rabbit, 1991), and even when word recognition is correct (Tun, McCoy, & Wingfield, 2009). There are numerous surveys that suggest many undesirable consequences resulting from elevated listening effort, including increased need for extensive recovery time after work (Nachtegaal et al., 2009), increased incidence of long-term sick leave from work (Kramer, Kapteyn, & Houtgast, 2006), early retirement (Danermark & Gellerstedt, 2004), and unemployment (Järvelin, Mäki-Torkko, Sorri, Rantakallio, 1997), as well as general disengagement from social activities (Grimby & Ringdahl, 2000). The topic of effort reduction thus holds social and economic importance in addition to its role in understanding basic auditory language processing. Specifically, the experience of people with hearing loss might be cause for exploring whether elevated effort results from some specific disruption in the predictive process.

Prediction and other higher level cognitive processes are thought to be especially important when the auditory signal is degraded, such as for people with hearing impairment, or for people in difficult listening conditions such as a noisy or reverberant room (Mattys et al., 2012; Pichora-Fuller et al., 1995), or when listening to a talker with an unfamiliar accent (Van Engen & Peelle, 2014). Any of these situations can degrade the speech signal, and it is hypothesized that listeners have to compensate by deploying extra cognitive resources to understand a spoken message (Hicks & Tharpe, 2002; Rönnberg, Rudner, Foo, & Lunner, 2008). This results not only in poorer recognition of words but also poorer memory for words that are heard correctly (McCoy et al., 2005; Rönnberg et al., 2013), suggesting that the extra load needed to overcome signal distortion interferes with other communicative or cognitive functions. If the speech were more predictable, perhaps some of the cognitive load of listening could be relieved. This has been verified by Sohoglu, Peelle, Carlyon, and Davis (2012) in an experiment that revealed reduced cortical activation when degraded (vocoded) words were preceded by relevant text. In the current study, we explore prediction entirely through the auditory channel, wherein the signal degradation could interfere with the process of prediction itself.

Pupillometry—the measurement of pupil dilation—was used in this study as a measure of listening effort. This metric has a long history as a general index of cognitive load (see Beatty, 1982; Laeng, Sirois, & Gredeback, 2012 for reviews), including various studies

of listening effort in response to masking noise (Zekveld, Kramer, & Festen, 2010) and signal distortion relating to spectral resolution and CI-like processing (Winn, Edwards, & Litovsky, 2015), as well as target-masker spatial orientation (Zekveld, Rudner, Kramer, Lyzenga, & Rönnerberg, 2014) and lexical competition (Wagner, Toffanin, & Başkent, 2015). The pupillary dilation response is driven by increased activity of the sympathetic branch (or *reduced* activity of the parasympathetic branch) of the autonomic nervous system, meaning it can be interpreted more broadly as an index of arousal or cognitive activity. The pupillary dilator muscles are thought to be driven by the locus ceruleus of the noradrenergic system (Wilhelm, Wilhelm, & Lüdtke, 1999), with which dilation has been shown to be phase locked (Aston-Jones & Cohen, 2005). A recent report by McGinley, David, and McCormick (2015) provides detailed analysis of other physiological measures that are correlated with pupil diameter in mice, including cortical membrane potential strength, theta activity in the hippocampus, eyelid opening, and spontaneous locomotor activity. They found that transient cortical depolarizations occurred roughly 1 s before pupil microdilations, consistent with other reports of a 1-s delay in peak pupil response in humans (Verney, Granholm, & Marshall, 2004). Although they used a simple tone-detection in noise task in mice, the presence of cortical (rather than simply low level) activity suggests that some of these physiological connections could underlie the task-evoked pupillary responses observed in numerous studies of human speech perception as well.

There has been little work done to test whether the task-evoked pupillary response is sensitive to the predictability of a stimulus. One published report by Qiuyuan, Richer, Wagoner, and Beatty (1985) demonstrated smaller pupil dilation in response to high-probability compared with low-probability non-speech stimuli. Their study had participants count the number of target and nontarget tones whose probability (frequency of occurrence) was controlled. Greater dilation was elicited by less-predictable tones or the absence of predictable tones. It is not clear whether this result translates directly to speech stimuli, where processing of content should be relatively more complex than the relative frequency of occurrence of target sounds.

Pupillometry caters to the specific goals of the current study in that it is an *ongoing* measure of effort whose temporal resolution is good enough to distinguish differences of effort within a sentence. In the experiment presented here, sentences differ in their presence or absence of semantic context that could be used to predict upcoming words. Although facilitation for word recognition can be measured with reaction time *following* the presentation of a target word, pupillometry offers a window into the processing leading up to that word, consistent

with the idea that processing does not begin only as a word is being uttered.

Relation to Theories of Speech Processing

Hearing impairment potentially alters the fundamental mechanisms of auditory language processing. Numerous theories of spoken word recognition postulate that ongoing input will activate matching candidates from a listener's lexicon, while words that don't match are suppressed or discarded (cf. the Cohort model of Marslen-Wilson & Welsh, 1978 and the TRACE model of McClelland & Elman, 1986). It was therefore speculated that when degraded signal quality renders a listener unsure about the reliability of incoming information, then the correct lexical possibilities would be less strongly activated, and the incorrect options would be improperly sustained. This could result in more cognitive resources engaged and sustained, because there is little to no input available to prune down the various lexical options and thus disengage cognitive activity. While this would not necessarily preclude the *correct* perception of the words, it could be a sign of increased cognitive cost, as the well-tuned lexical activation mechanisms would not effectively constrain the expected input as rapidly and efficiently as the process described by Lau, Stroud, Plesch, and Phillips (2006), whose evoked potential results suggest constraint of word categories as early as 200 ms post stimulus in response to clear quiet speech. The results of Kuchinsky et al. (2013) suggest that increased lexical activation elicits greater effort, particularly when signal quality is degraded. In that study, greater pupil dilation (i.e., greater effort) was elicited for words where phonological neighbors were explicitly offered as response options. The current study uses a format where no options are explicitly presented, but activation of multiple lexical competitors is hypothesized because of the signal degradation that would arise with the use of a CI or noise vocoder.

The Goals of the Current Study

This study set out to examine (a) whether listening effort is reduced by the opportunity to predict spoken language on the basis of semantic context, (b) how quickly effort reduction (if any) can be observed, and (c) whether prediction-related reduction of listening effort is contingent on the quality of the sound input/auditory system.

It was hypothesized that listeners with normal hearing (NH) would show elevated pupillary responses when listening to vocoded speech, (consistent with the results of Winn et al., 2015) and that this elevation would also be observed in listeners with CIs on account of their degraded auditory input. Furthermore, it was expected that high-context sentences would elicit a smaller

pupillary response than low-context sentences. We expected that previous findings of deficits in efficient predictive processing in some populations (Federmeier, McLennan, & Ochoa, 2002) would emerge for CI listeners and for the NH listeners in the vocoder condition on account of the lack of clear signal quality. The operational measure of predictive processing was a reduction in pupillary response for high-context versus low-context sentences, as described below in the Methods section.

Methods

Participants

Participants included 20 listeners with NH who heard regular speech and distorted (“vocalized”) speech, as well as 12 listeners with CIs, who were also older (see Table 1). All were native speakers of American English and reported no learning difficulties or cognitive impairments at the time of testing. Two additional NH listeners were tested but ultimately excluded from data analysis due to testing difficulties relating to pupillometry (e.g., poor tracking and inattention). Nearly all CI listeners were postlingually deafened, but age of onset of deafness was not an exclusion criterion. All experimental protocols were approved by the Institutional Review Board of the University of Wisconsin–Madison, and all listeners provided informed written consent prior to participation. Listeners were not screened for the use of drugs that could affect physiological responses. Although none of the experimenters involved in the study has training in detecting the influence of drugs, no participants gave any overt signs of atypical behavior that would suggest a chemically altered state of mind.

Stimuli

The Revised Speech in Noise (Bilger et al., 1984) stimulus corpus was used for this experiment. This corpus contains lists of 50 sentences each; each list contains 25 sentences where a final target word is preceded by semantic context, resulting in “high-probability” target words (e.g., “Stir your coffee with a spoon”), with the other 25 sentences lacking any context that would help predict the target word (e.g., “Jane thought about a spoon”). High-context and low-context sentences were mixed within each list and they were not cued in any way.

Stimuli were presented to NH listeners in two ways. They were presented as full-spectrum normal speech, and also presented with spectral degradation in the form of an eight-channel noise vocoder, which is commonly used to match the speech intelligibility performance of better-performing CI listeners (Friesen, Shannon, Başkent, & Wang, 2001). The vocoder processing extracted eight spectral bands between 150 and 8,000 Hz that were equal in terms of cochlear spacing (estimated using the formula provided by Greenwood, 1990); the temporal envelope of each band was extracted and used to modulate a corresponding band of noise matched in frequency bandwidth. The envelopes were low-pass filtered with a 300-Hz cutoff frequency, which was sufficient to encode the fundamental frequency of the talker. Following the modulation of the eight channels, all bands were summed to create a final stimulus whose intensity was matched to that of the original unprocessed sound, but whose spectral detail was considerably impoverished. CI listeners only heard unprocessed sounds. All stimuli were presented in quiet.

Table 1. Demographics of Cochlear Implant Participants.

CI Participant	Sex	Age	Age of onset of HL	Device R	Device L	Years of CI exp.	Years of BiCI exp.
1	F	67	13	Nucleus 5	Nucleus 5	16	9
2	M	66.7	birth	Freedom	Nucleus 24	16.5	9.5
3	F	69.3	12	Freedom	Nucleus 24	14.5	9.8
4	F	63	13	Nucleus 5	Nucleus 5	3	1
5	M	40	4	Freedom	Nucleus 5	6.8	2.8
6	M	74	51	Nucleus 5	Nucleus 6	11	5
7	M	62.6	unknown	Freedom	Freedom	0.33	0.1
8	M	20.8	7	Freedom	Nucleus 5	6.25	3.25
9	F	48.7	5	Freedom	Freedom	2.25	1
10	M	71.3	10	90 K HiRes	90 K HiRes	17.5	6.4
11	F	67.2	21	Hearing aid (off during testing)	Nucleus 5	1.75	NA
12	F	52.4	22	Hearing aid (off during testing)	Nucleus 5	1.7	NA

Note. CI = cochlear implant; HL = hearing loss; BiCI = bilateral cochlear implant; NA = not applicable.

Procedure

Listeners were first familiarized with the testing setup in terms of the target talker's voice and the pace of stimulus presentation during a practice block of 6 to 10 sentences. Participants with NH then heard another 6 to 10 sentences of vocoded stimuli; none of the practice stimuli were repeated during testing. Listeners sat in a comfortable chair in a sound-treated room in front of a computer screen that was 4 feet away; the screen occasionally showed written instructions but primarily had a simple red cross for the listeners to visually fixate on during testing. Luminance of the screen and the testing area were kept constant throughout the entirety of the testing session, with the screen filled with the gray color defined as [102 102 102] in the RGB color space. Participants' heads were stabilized using a chin rest (SR Research, Inc., Mississauga, Ontario, Canada), which facilitated reliable pupil size measurements. Pupil diameter was recorded using an SR Research EyeLink 1000 Plus using 500 Hz sampling frequency in binocular tracking mode.

During testing, each sentence was preceded by 2 s of silence to permit measurement of baseline pupil diameter. Stimuli were presented at 65 dBA through a single Tannoy loudspeaker located in front of the listener. Each sentence was followed by 2 s of silence, after which the red fixation cross changed to a green color of equal luminance to signal that the response should be given. All participants reliably detected the color change. After each verbal response from the participant, the experimenter scored the accuracy of reporting the target word as well as the preceding portion of the sentence (i.e., the "context"). Substitution of function words (e.g., "the" for "a") were not penalized. Following participant responses and hand scoring, an additional 4 s of silence followed each trial, to allow the pupil to return to baseline and stabilize. Occasionally, participants' pupil sizes would adapt to the local luminance, and they were given short breaks so that their pupils would return to a state where changes in size could be detected.

Testing was divided into four blocks. First, a block of 25 unprocessed sentences was presented, followed by 25 vocoded sentences from another list. Then sentences 26 through 50 from the first unprocessed list were finished and then sentences 26 through 50 of the vocoded list. CI listeners heard just three blocks of 25 unprocessed sentences; more stimuli were used for CI listeners because, in light of the relative difficulty in recruiting the clinical population, the potential exclusion of participants due to insufficient data was intended to be minimized. Each block contained a break after 13 sentences to give the participant a chance to rest and look away from the screen.

Data Processing

Pupil dilation data were subject to a multistage data cleanup and transformation process that was an extension of procedure reported by Winn et al. (2015). Stretches of missing data corresponding to blinks were expanded asymmetrically such that 40 samples (80 ms) prior to and 70 samples (140 ms) following a missing data stretch were omitted. This was motivated by the procedure described by Zekveld et al. (2010) and resulted in the exclusion of local pupil size disturbances caused by the motion of the eyelids into and out of a blink. Following the exclusion of blinks, data were linearly interpolated across stretches of missing samples. A 10-Hz low-pass filter was applied to smooth the data. For each trial, mean pupil size was calculated during the 1-s period preceding each stimulus to establish baseline, from which all subsequent measurements were compared.

Individual trial data of pupil dilation over time were visually inspected to check for any residual mistracking following data cleanup. Primarily, this was done to identify baseline periods where there was a gross excursion from resting state, which would have contaminated all subsequent data from that trial, on account of the convention of referencing to baseline. Trials that contained substantial amounts (~40%) of missing data, gross artifacts, or missing data specifically within the general region of interest (proximal to the delay between stimulus and response) were discarded, following the procedure described by Verney et al. (2004). Data rejection resulted in the exclusion of 20% of all trials for CI listeners and NH listeners in the vocoder condition and the exclusion of 27% of trials for NH listeners in the unprocessed condition (more data loss was expected for the unprocessed condition on account of the general lack of task engagement needed to complete this easy task). Following this process, the trial-level time-series data were aligned to stimulus offset and were aggregated by condition (unprocessed/vocoded), per context type.

Pupillometry data were quantified by measuring proportional change in pupil diameter relative to the baseline period before each stimulus. Baseline was defined as the 1-s period preceding stimulus onset. A minimum of 50 ms (25 samples) was required for a valid baseline estimation; in the case of prolonged blinks during the baseline period, the baseline window extended backward to include more samples until a minimum of 25 samples were obtained.

Analysis

Intelligibility errors were divided into two categories: *context* and *target*. Errors on the final word were counted as target errors; any other errors (on words

earlier in the sentence) were considered as context errors, even in the case of low-probability sentences that contained no *useful* context. Intelligibility analysis was concerned with the proportion of sentences in each condition that contained a context error or a target error, as well as target errors that were specifically preceded by context errors.

Growth curve analysis (Mirman, 2014) was used to model change in pupil dilation over time, as previously done by Kuchinsky et al. (2013) and Winn et al. (2015). This analysis consists of using summed orthogonal polynomials to model changes in overall level, slope, and inflection of a continuous variable over time. The simple linear combination of polynomial components permits the use of hierarchical (i.e., mixed-effects) analysis, with dynamic consistency across nested levels (Mirman, 2014).

Pupil dilation was modeled as a function of two explicitly planned predictors: hearing group (NH, NH vocoded, or CI) and sentence context (coded as binary high/low) as they interacted with time. This approach modeled the change of effects over time rather than static effects at a particular time point or from a summation of time points in a bin. The statistical model utilized a second-order orthogonal polynomial, with interactions between hearing group, context, and each of the three orthogonal time components (intercept, linear, and quadratic). This means that for every combination of hearing group and context level, the three descriptors of the response curves (height, slope, and inflection) were independently estimated and compared for significant changes against a default condition. Intercept (overall level) was used to estimate *overall* level of pupil dilation and is akin to a basic analysis of variance analysis; the linear (slope) term was used to gauge growth in pupil dilation over time. The quadratic component was used to model the shallowing of the growth function as it reached its peak and primarily was used to improve model fit rather than to correspond to a specific prediction. The context factor was used for the hypothesis that high-context sentences would elicit smaller overall pupil dilation (Intercept), as well as shallower growth (linear slope) of pupil dilation. Each of the model terms were used in a generalized linear mixed-effects model fit using a maximum likelihood estimation procedure using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R software environment. Model selection was performed using a constrained backward stepwise elimination procedure, starting with a model that contained all potential terms and interactions (hearing group and context). The polynomial time (intercept, linear, and quadratic) terms never interacted, on account of their deliberate orthogonal nature. The initial model estimation was made using a third-order polynomial, inspired by the apparent existence of

two inflections in the data. However, a model with no third-order (cubic) polynomials was deemed to be significantly more parsimonious, based on a Chi-square test ($p < .01$), and a reduction in the Akaike Information Criterion, which is a diagnostic of goodness-of-fit that is penalized by model complexity. The prevailing model took the following form:

$$\begin{aligned} &\text{Pupil dilation} \\ &= \text{poly1} \times \text{Group} \times \text{Context} + \text{poly2} \\ &\quad \times \text{Group} \times \text{Context} \\ &\quad + (1 + \text{poly1} + \text{poly2} + \text{Context} | \text{Listener}) \\ &\quad + (1 + \text{poly1} + \text{poly2} + \text{Context} | \text{Group}) \end{aligned}$$

where the left side of the = is predicted by the terms on the right side; “×” refers to an interaction between model terms, including all main effects, two-way and three-way interactions. Poly1 and poly2 are orthogonal linear and quadratic time terms, respectively. Terms in parentheses are random effects, which are estimated to vary within a distribution that partially explains the variation in the fixed variables of interest. There were subject-level random effects of intercept (“1”), linear and quadratic polynomial time, and context, meaning that all of these factors were free to vary by subject, centered on the group mean. There were also group-level random effects of intercept, linear and quadratic time, and context, meaning that these factors varied across groups, within which the magnitude of the factor coefficients was estimated to vary within a normal distribution.

In the model described earlier, hearing *Group* consisted of three groups—NH (unprocessed), NH (vocoded), and CI listeners. It should be observed that the data from NH listeners in both conditions are not independent; data from a single NH listener in the full-spectrum/unprocessed condition are related to data from the same NH listener in the vocoded condition. Inclusion of that dependency in the unified model would have required an extensive number of undesirable three-way and four-way interactions that could obfuscate the interpretation, and also not be applicable to CI listener, who only heard one *condition*. Thus, although the dependence of data across conditions is a real part of this data set, it was not explicitly modeled for the current analysis.

Consistent with the approach of Verney et al. (2004), responses were broken into multiple time windows, in this case to separately reflect the (a) *listening* and the (b) *rehearsal* portions the trials, after adjusting for the delay of the pupillary response. The listening (first) window began 1 s prior to stimulus offset and ended 1 s after the stimulus, and the rehearsal (second) second window began at the offset of the first window and continued to 1 s past the response prompt (3 s poststimulus offset). These timeline landmarks were chosen because

task-evoked pupillary responses have latency between 0.7 and 1.5 s (Verney et al., 2004), meaning that for each auditory event, the corresponding pupil response will be shifted in time. The time windows and their corresponding pieces of the trial timeline are illustrated in Figure 1.

A second analysis was carried out upon the observation that the reduction of effort associated with context is quantified directly in the time course of area *between* curves for high- and low-context sentences. Differences between curves were measured within the time window that was slightly wider than that used for the previous analysis; it spanned -3 s to $+4$ s relative to stimulus offset, which essentially captured enough time before the pupil dilation so that all data would begin at zero (as there should theoretically be no difference between high- and low-context sentences *before* the pupil dilates) and continued through the middle of the verbal response, in order to capture late-occurring effects. Differences between curves took the form of a traditional sigmoid shape and were thus estimated using a conventional nonlinear least squares (NLS) approach. A three-parameter sigmoid was fit to the data, including free parameters for the slope, upper asymptote, and x -axis shift (intercept). Lower asymptote was fixed at zero, consistent with the expectations explained earlier. The prevailing model was defined abstractly as follows:

$$\text{Curve} = \frac{\text{upperasymptote}}{(1 + \exp(-1 \times \text{Time} \times (\text{slope} \times \text{shift})))}$$

and took the following specific form for this study:

$$\begin{aligned} \text{Percent reduction} \\ = (\max \times \text{group}) / (1 + \exp(-1 \times (\text{Time} + \text{shift} \\ \times \text{group}) \times (\text{slope} \times \text{group}))) \end{aligned}$$

This sigmoidal analysis complements the growth curve analysis in that both can demonstrate differences between conditions for the various listener groups. The added value of the sigmoidal analysis is that (a) it is simpler to conceptualize, compared with orthogonal polynomials and (b) it allows the estimation of three potentially very meaningful parameters, including the overall amount of difference and the latency of a specific amount of difference (defined here as effort release). In question was the time at which a specific growth away from zero was observed. Lacking any precedence for such a measurement, 10% reduction (with reference to the maximum dilation observed in the low-context condition) served as threshold to define the excursion. The earliest time that elicited this threshold was defined as the latency of effort release. Results of this model and latency of effort release are displayed in Table 2 (described further in the Results section).

Results

Intelligibility

Listeners with NH had virtually perfect intelligibility scores for regular full-spectrum speech sounds regardless

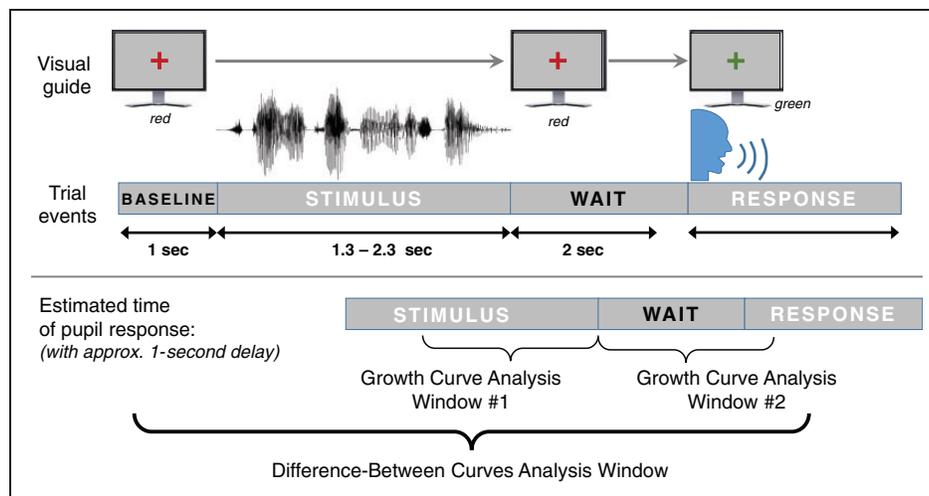


Figure 1. Timeline of events during experimental trials. The monitor seen by the participant contained a colored fixation cross that changed to green to elicit a verbal response. The *trial events* show the timeline of the auditory stimuli, while the timeline of the pupillary responses are estimated to arise roughly 1 s after their corresponding auditory events. Windows for growth curve analysis are indicated by two small brackets that include the stimulus, and the wait time, respectively. The window for difference-between curves analysis is represented by the large bracket that included the entire stimulus, wait period, and part of the verbal response.

of the presence or absence of context. When those listeners heard vocoded speech, intelligibility was considerably poorer; 11% of high-context and 38% of low-context sentence responses contained at least one incorrect word. Among all mistakes on final (“target”) words in high-context sentences, only 9% were preceded by an error on a *context* word, suggesting that misperception of context alone was not the driving force behind target word errors. The corresponding sentence error rates for listeners with CIs were 7% and 33% for high- and low-context sentences, with just 5% of high-context target word errors preceded by an error in reporting the context.

Table 2. Parameters of a Sigmoidal Model Estimating the Percentage of Reduction in Pupil Dilation Attributable to Semantic Context.

Group	Maximum (%)	Slope	Shift	Latency (s)
Normal hearing	32.0	1.20	−0.13	−0.79
Normal hearing vocoded	14.2	4.60	1.75	1.94
Cochlear implant	23.4	1.03	0.57	0.29

Note. The last column reflects latency (seconds relative to stimulus offset) to 10% reduction of peak dilation, and is distinct from “shift” which refers to shift of the function midpoint relative to the center of the analysis window.

Pupillometry

For listeners with NH, overall pupil dilation was elevated for vocoded speech, consistent with earlier work (Winn et al., 2015). Pupillary responses from listeners with CIs were lower than those for vocoded speech, but higher than those for NH listeners for unprocessed speech. Noteworthy differences in the pupillary response were observed by comparing high-context to low-context sentences, as described later.

For listeners with NH, full-spectrum/unprocessed sentences with high context showed reduced pupil dilation compared to those with low context, as can be seen in Figure 2 (left panel). On average, that reduction *emerged before the end of the sentence*, suggesting that processing of semantic context resulted in rapid effort release during the perceptual process; the reduction continued through to the verbal response portion of the trials. In contrast, when listening to vocoded speech—signals with degraded spectral quality—semantic context did not yield a reduction in pupil dilation *until after the sentence was over*, suggesting that listeners were not able to rapidly exploit the incoming information to reduce effort. Listeners with CIs showed some context-related reduction in pupil dilation that was statistically detectable (described later) but reduced compared with what was observed in normal-hearing listeners.

For NH listeners in the normal speech condition, growth curve analysis revealed significant differences

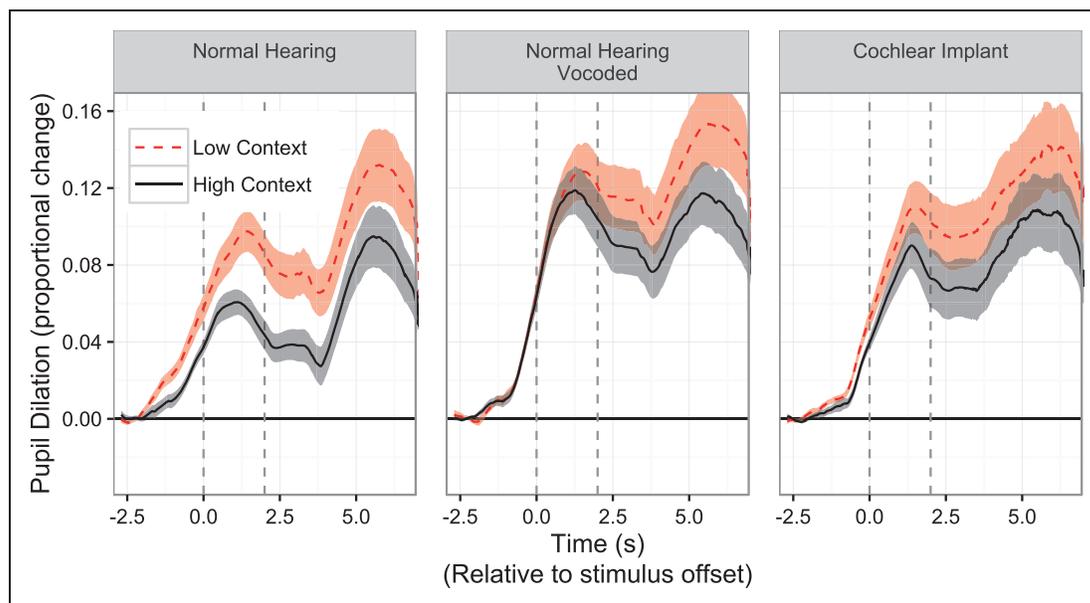


Figure 2. Relative change in pupil dilation in response to two sentence types, observed in listeners with normal hearing (left and center panels) and listeners with cochlear implants (right panel). Greater magnitudes correspond to greater listening effort. The left dashed line (Time 0) shows the offset of the sentence while the right dashed line (Time 2 s) shows the timing of the response prompt. “High context” sentences contain early-occurring words that are semantically related to later-occurring words, allowing for prediction. In such cases, a reduced pupillary response is observed, particularly for listeners with normal hearing listening to normal speech (left panel).

between high- and low-context sentences ($p < .01$ for the overall level, slope, and inflection components of the curves), which were smaller but still statistically significant ($p < .01$ for each component) for participants who used CIs. For the vocoder condition heard by people with NH, no statistically detectable difference in curves was observed in the first time window ($p = .79; .17; .34$ for the effects of context on the intercept, linear, and quadratic components, respectively). In other words, no evidence of early occurring context benefit was found in the vocoder condition.

In the second analysis window, significant differences in slope were observed between the high- and low-context conditions for all three listener groups. Differences in intercept essentially reflected differences in the ending point from Window 1; there was a significant difference between intercepts for the NH unprocessed and NH vocoder condition ($p < .01$), but no statistical difference between the vocoder condition and the intercept for the CI listeners ($p = .44$). In the second window, the slope parameter was negative, indicating a return toward baseline following the peak at the end of Window 1. For NH listeners, there was a marginally greater magnitude of the negative slope in the full-spectrum condition compared with that obtained in the vocoder condition ($p = .09$) and a marginally larger magnitude of negative slope for the CI group compared with the NH full-spectrum group ($p = .10$); there were no differences between the vocoder group and the CI group.

Of particular interest is the difference *between* curves in Figure 2, which is interpreted as the difference in effort related to the predictability of the sentences. In parallel with other factors that improve intelligibility, an

appropriate term for this phenomenon would be *effort release*, resulting from predictions based on semantic context. This concept is directly visualized in Figure 3.

Effort release was calculated using a standard non-linear least squares fitting procedure with a three-parameter sigmoidal formula described in the Methods section. Parameters of this model are listed in Table 2. Latency to reach 10% reduction of pupil dilation (compared with the low-context condition) was estimated using the model predictions. Effort release was largest and earliest (1 s before stimulus offset) for listeners with NH who heard full-spectrum (non-vocoded) speech sounds. In conjunction with the relatively slow time course of the pupillary response, this is conclusive evidence of context benefit *during* the perceptual process. Effort release was considerably delayed—by over 2.5 s—for the same listeners when hearing vocoded speech, suggesting that processing occurred *after* (or at least at the *end* of) the perceptual process. Latency to effort release was slightly delayed—by roughly 1 s—for the group of listeners with CIs, suggesting a slight delay in processing, but not as slow as that observed for NH listeners in the vocoder condition. Latency of effort release for CI listeners likely indicates some early-occurring benefit of context, despite listening to a relatively degraded signal.

Validation of Statistical Models

To verify that the statistical models provided a good estimation of the data, the predictions of the models were compared to the actual group data. Figure 4 illustrates the model prediction of the pupillary responses

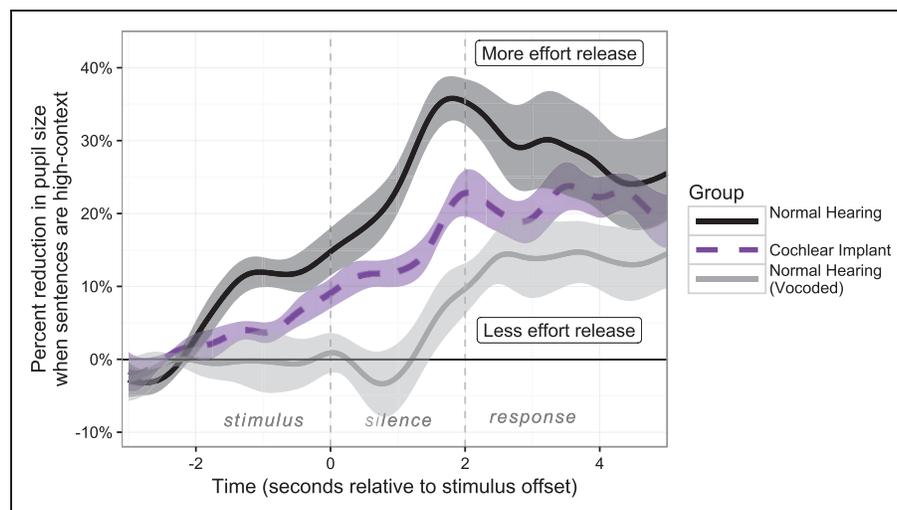


Figure 3. Differences between curves illustrated in Figure 2, expressed as percentage reduction of pupil size compared with low-context condition. Greater magnitudes indicate more benefit from context as indicated by greater reduction in pupillary response. The left dashed line (Time 0) shows the offset of the sentence while the right dashed line (Time 2 s) shows the timing of the response prompt.

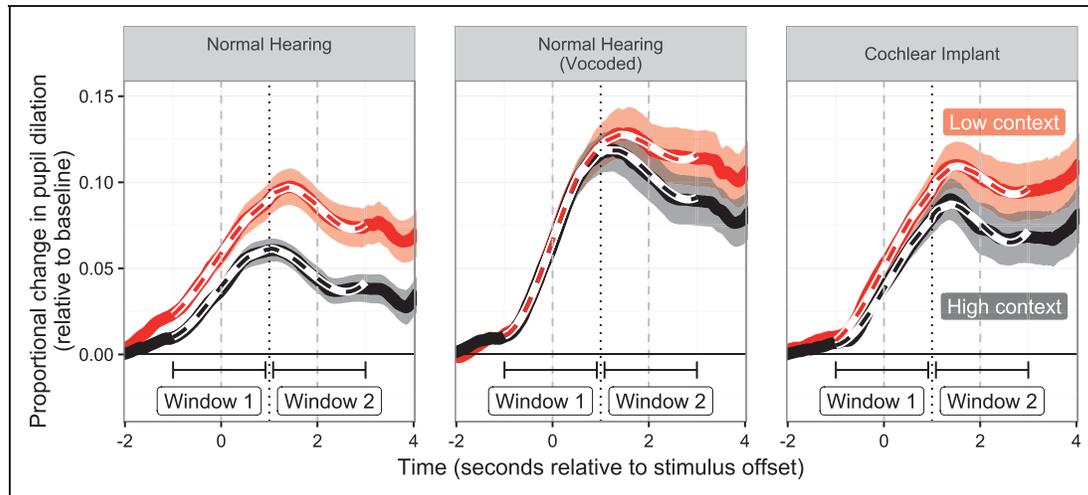


Figure 4. Validation of the polynomial growth curve model used to estimate pupillometric data displayed in Figure 2. Model fits (dashed lines) are overlaid on ribbons that reflect ± 1 standard error of the mean of the group data. Separate models were used to estimate Window 1 (reflecting the perception phase of the trials) and Window 2 (reflecting the planning and response phase of the trials).

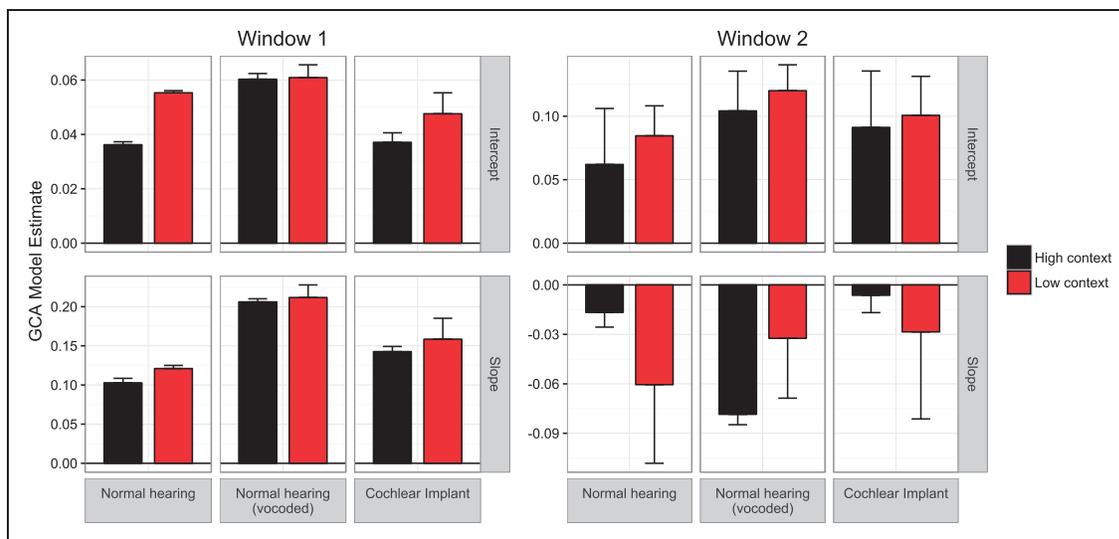


Figure 5. Illustration of the intercept slope parameter of the polynomial growth curve models displayed in Figure 4. Intercept reflects overall level while slope reflects the rate of pupillary growth during the indicated time window. For Window 2, negative slope corresponds to the rate at which pupil size returned to baseline.

previously shown in Figure 2, for both time windows in all three listener groups. Predicted data are dashed lines overlaid on observed data standard-error ribbons. It can be seen that the polynomial fits accurately reflected the pattern observed in the observed data.

The intercept and slope parameters—the model terms that reflect overall level and rate of growth in pupil dilation, respectively—are arguably the most important for interpreting these data. Figure 5 illustrates the magnitude of the intercept and slope parameters for all listener groups, for both high- and low-context conditions.

During Window 1 (the *perception* part of the trial), the slope reflects the rate of pupillary response growth, which is lowest for NH listeners, highest for NH listeners in the vocoder condition, and intermediate for CI listeners. There were significant differences in slope related to the level of predictability in the sentence materials for the NH and CI groups, but not the NH vocoded group. During Window 2, slopes are negative because pupil size decreases after the response to the stimulus. As reported previously, the context-related differences for slopes in Window 2 were significant for all three groups,

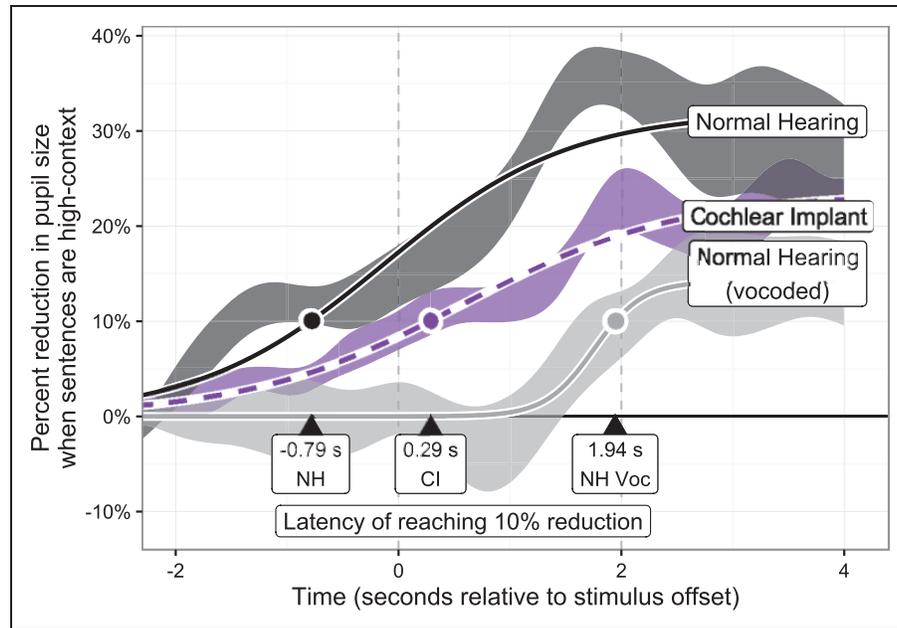


Figure 6. Validation of the three-parameter sigmoidal model used to estimate differences between curves displayed in Figure 3, that is, the effect of context on the reduction of pupil dilation. Model fits (dashed lines) are overlaid on ribbons that reflect ± 1 standard error of the mean of the group data. Dots reflect latencies (directly labeled) for the predicted curves to reach the level of 10% reduction relative to the maximum value observed for the low-context sentences.

suggesting that context has persistent effects on listening effort, even if those effects arise late (as in the case of NH listeners in the vocoder condition).

Figure 6 illustrates the prediction of the model of the sigmoidal shape of the difference between curves (i.e., “effort release”) in the observed data driven by the effect of context/predictability. Predicted data are dashed lines overlaid on observed data standard-error ribbons. Also shown are the calculated latencies for the predicted curves to reach the level of 10% reduction relative to the maximum value observed for the low-context sentences. Although the fitted data estimated by the sigmoidal model are not as accurate as the models derived from the growth curve analysis, they provide a clean snapshot of the differences between curves within a more conventional framework that can be easily applied to other pairs of time-series data.

The influence of sentence intelligibility was examined in conjunction with pupillary responses so that the difference in effort between conditions (or lack thereof) could not be explained merely by differences in performance accuracy. Figure 7 illustrates pupillary responses for trials where intelligibility was perfect overall, (upper panels), and for trials where intelligibility was perfect for all of the context words leading up to the target (predictable) word (lower panels). It can be seen that the main findings of context-related differences in pupillary responses in NH listeners and CI listeners, and the lack of such an early onset in NH listeners in the vocoder

condition, are maintained even for sentences with no intelligibility errors. These results suggest that intelligibility scores, while sensitive to context benefit, are not able to reveal how degraded listening conditions could give rise to differences in the time course of listening effort. In other words, even sentences that are reported correctly could be processed without rapid benefit of context.

Discussion

This study examined the impact of context on speech intelligibility and ongoing listening effort (reflected by change in pupil dilation over time). Intelligibility measures were consistent with previous literature that suggested an advantage for words preceded by relevant semantic context. Pupillometric measures in this study suggested that this advantage can also be seen as a reduction in listening effort. Following previous literature on factors that influence masking, this effect has been termed here as effort release. Effort release stemming from context occurs rapidly for people with NH and is slightly delayed for people who use CIs. NH participants listening to vocoded speech showed substantial delays in effort release, implicating a disruptive role of spectral resolution (i.e., signal clarity) in one’s ability to exploit context efficiently. Although CI listeners experience poor spectral resolution, they also have considerably more experience with their devices. It is reasonable to suspect

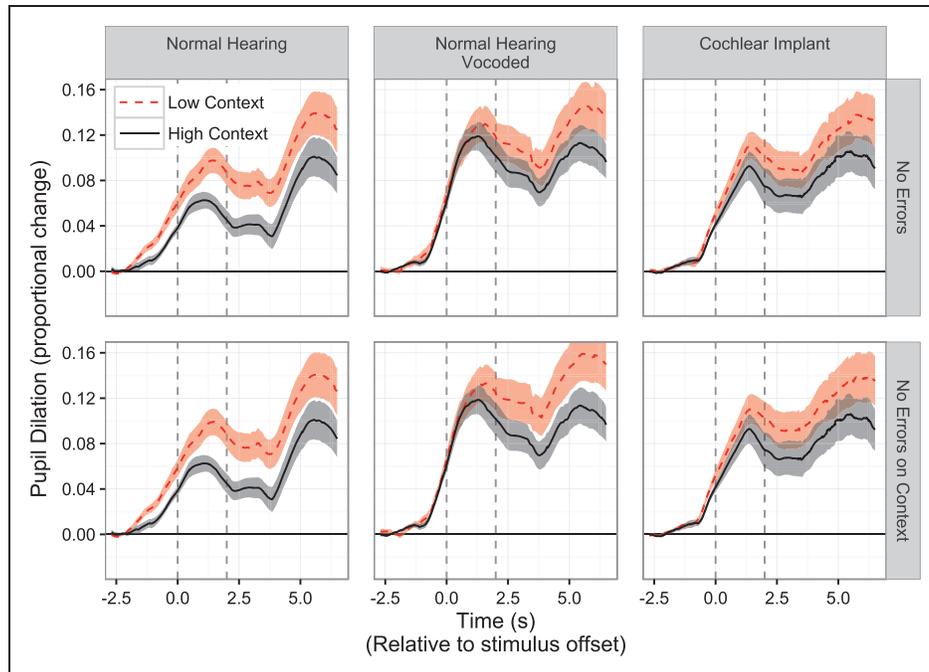


Figure 7. A re-illustration of data displayed in Figure 2, including only trials in which participants' verbal responses contained no errors on all words leading up to the final (target) word (lower panel) or no errors at all (upper panel). The preservation of the general pattern observed in Figure 2 suggests that context-driven differences in listening effort across time can emerge even when intelligibility is high. The left dashed line (Time 0) shows the offset of the sentence while the right dashed line (Time 2 s) shows the timing of the response prompt.

that they have gained the ability to handle distorted/unclear speech in ways that are unpracticed in nonimplanted listeners. However, the possibility also exists that the CI listeners and NH listeners are equally adept in their ability to benefit from context, but demonstrate differences due to other factors, such as age or general cognitive abilities.

The potential effects of age and fatigue are noteworthy for this study. Winn, Whitaker, Elliott, and Phillips (1994) have shown that older individuals show generally less pupil dilation than their younger counterparts. In their study, pupil dilation in response to different luminance levels was measured across a wide age range, yielding a linear effect of age (but no effect of gender or iris color) that was most pronounced at lower luminance levels. In the current study, the CI listener group was substantially older than the NH control group, opening up the possibility that reduced pupil size for the CI group compared to the NH vocoder group could potentially have resulted because of age differences. In other words, it could be the case that the CI listeners would have shown just as much elevated pupil dilation as seen in the NH vocoder condition if the participants were age matched. Additionally, even just a few minutes of fatigue can also restrict the range of task-evoked pupillary responses (Lowenstein & Loewenfeld, 1964), leading Hess (1972) to caution against the presentation of a large number of stimuli in pupillometry

experiments. In light of the known elevated levels of listening effort for people with hearing loss, these cautions could be especially important for investigations of effort in clinical populations.

The current results suggest that the benefits of context are not always captured in intelligibility scores. Specifically, the benefit can be rapid or late, but these two options are indistinguishable from basic intelligibility scores, despite being potentially very meaningful to the listening experience. It is feasible to speculate that good intelligibility can be the result of reflection/perceptual restoration after a sentence is heard, rather than immediate recognition of all words. In light of the common convention of stimulus-then-silence style of speech perception testing, it is possible that difficulties involving processing speed are not captured by most clinical and experimental tasks. Because of the open-ended nature of the testing paradigm, participants could consider each sentence as a whole to produce well-formed responses despite misperceiving individual words (e.g., "The bird was made from whole wheat" will be reported *correctly* as "The bread was made from whole wheat," despite a mistake in hearing "bread" as "bird"). Sentence recognition that unfolds due to correct perceptions and efficient prediction is likely a vastly different experience than sentence perception due to partial correct perceptions supplemented by retroactive perceptual restoration. Clinically, the value of this "correct-only"

sentence analysis is that it shows that even when accuracy is high, there could be differences in processing speed or processing strategy (e.g., reliance on top-down processing) that could affect communication success, but which are not captured by accuracy scores. By observing cognitive activity *throughout* the perceptual process, we can distinguish these perceptual phenomena.

The disruption of the unfolding process of speech perception due to signal distortion has been prescribed as a necessary part of any spoken word recognition model (McQueen & Huettig, 2012). Typically, this disruption has concerned the *timing* and *accuracy* of lexical recognition but has also been described in terms of lexical prediction bottlenecked by higher level constructive processes (Chow, 2013). In this study, we further support this latter notion by demonstrating a delay in the ability to exploit context to reduce effort when the speech signal is distorted. The current study was not designed to distinguish between any particular models of spoken word recognition (i.e., there were no clear opposing hypotheses motivated by different theories), so its role in the theoretical arena is limited. However, the consistency of the sustained pupillary response with the idea of sustained lexical activation could be an area for future research aimed more specifically at effort resulting from lexical factors, in a manner consistent with the study by Kuchinsky et al. (2013).

The observation of clear effects *during* the perceptual process for normal-hearing listeners supports the rapid predictive-processing framework of speech and language perception, such as the interactive lexical activation model described by McClelland, Mirman, Bolger, and Khaitan (2014), as well as the general linguistic-cognitive framework described by Lupyan and Clark (2015). This perspective has previously been supported by image-focused gaze tracking, reading tracking, and cortical-evoked potentials that demonstrate the ubiquity of prediction in language processing. Reduction in autonomic arousal for semantically predictable sentences in this study is consistent with the reduction in activity in left anterior temporal cortex measured by Lau et al. (2013). However, effects in this study stretched over a relatively long time frame that was even longer than that recently described by Lau and Nguyen (2015). Pupillary responses to difficult listening conditions measured by Winn et al. (2015) also revealed sustained effects in situations where intelligibility was below 100%. In that study and in the current study, the sustained elevated effort driven by degraded speech appears to carry forward through the time and affect the cognitive load involved in the verbal response (particularly for low-context degraded sentences in the current study).

The relatively long-lasting effects of context in this study imply continued cognitive processing during the silent time allotted for verbal response, which might

not be available during normal flowing conversation. Pauses in conversational speech (by a single talker continuing to speak), if present at all, are estimated to be between 300 and 730 ms, depending on the study (cf. Heldner & Edlund, 2010, and references therein). An analysis of intertalker transitions (i.e., conversation turn-taking) by Heldner and Edlund (2010) suggests that transitions between utterances are not only reliably shorter than 1 s, but frequently involve *overlap* of speech, such that one sentence begins before the previous one has ended. It should be noted, however, that talkers generally tend to adapt their turn-taking (i.e., speech gap/speech overlap) behaviors with respect to their conversational partners (ten Bosch et al., 2004).

Pupillometry can potentially play a very specific role in the exploration of listening effort for people with CIs. Like other physiological measures, it has the advantage of objectivity, unlike studies of self-reported effort, which can be impaired by interpersonal differences in willingness to admit effort, as well as more general demographic trends in the reliability of reporting effort (Kamil, Genther, & Lin, 2015). It has recently been hypothesized that self-report measures of effort and pupillometry simply represent different *kinds* of effort (Wendt, Dau, & Hjortkjaer, 2016; Zekveld & Kramer, 2014), including cognitive overload. While the temporal resolution of pupillometry is far more crude than electro/magnetoencephalography (EEG/MEG), the electronics involved in CIs provide considerable barriers to using those technologies with this clinical population. Functional near-infrared spectroscopy is also compatible with the implant, although its temporal resolution at the time of this writing is not fine enough to discern changes in effort at the level discussed in this article—the use of semantic context within a sentence.

Behavioral measurements like reaction time or dual-task interference (cf. Hornsby, 2013; Pals, Sarampalis, & Başkent, 2013) are also used as objective measures of effort, but differ from pupillometry in that they are arguably limited to single-time point measures. Additionally, participants could vary in their ability to perform a secondary task or to multitask in general, which could add unwanted variability to a dual-task index of effort. Unlike self-reported effort or measures of reaction time, pupillometric measures are time-series data that can track the elevation of effort *during* a prolonged perceptual event such as a sentence and subsequent verbal response. The specialized equipment, expense, and fussiness of pupil data collection (and the complexity of multiple contributing influences to pupil size) render pupillometry an impractical tool for individual clinical practice at this time, but the method could potentially serve to indicate or contraindicate broad courses of treatment at the group level. For example, the use of particular speech-processing strategies or electrode

configurations in CIs might reliably elicit less effort for groups of people, which could guide clinical decisions in the same way that they could be guided by any other population studies.

Nearly all outcome measures for people with CIs are highly variable across individuals, presumably because of the confluence of numerous factors, including duration of deafness, etiology, surgical placement, and so on (Lazard et al., 2012), in addition to individual differences in cognitive and linguistic capacity. However, one commonality among all CI users is that they experience spectral degradation on account of the nature of their hearing device. As such, fast and predictive linguistic processing is jeopardized in this population (and potentially the larger population of nonimplanted people with hearing loss).

The late-occurring benefit in the degraded conditions in this study is consistent with the oft-experienced phenomenon wherein a listener in a noisy room asks “what?” after mishearing speech, only to quickly figure out what was said even before the conversation partner repeats the message. The substantial reduction of pupil dilation *after* the conclusion of high-context sentences (observed in Figure 2 center and right panels) might correspond to such a recovery process. The current results imply that that this post hoc restoration of sentences might be more common in listeners with hearing impairment than previously known (as intelligibility could be maintained at a high level via higher level cognitive processes despite misperceptions of the speech signal). Importantly, this phenomenon can be described using time-sampled measurements such as pupillometry. At this time, it is not clear how pupillometric results can directly inform the *alleviation* of listening effort, but the current study suggests that the speed and temporal separation of sentences might offer the chance for a listener to *catch up* and not fall far behind. As much of the benefit of context (reduced pupil size) was observed after the sentence, one might conclude that that’s when the context is being used to recover the lexical items in the sentence. If that time were kept protected by silence, a listener might enjoy greater success. This hypothesis was not directly tested in this experiment, but can feasibly be addressed by follow-up work, motivated by the observation of increased difficulty among older adults and adults with hearing loss when speech rate is faster (Wingfield et al., 2006).

The effects illustrated in this study highlight a need to reevaluate the abilities measured by conventional audiological and laboratory tests of word and sentence recognition, where speech is presented and then followed by an open-ended quiet time to give a response. While a keen observer (e.g., an experienced clinician) can detect signs of uncertainty in a response, it is clear that at least two relevant pieces of the speech perception process can be misestimated or discarded completely. First, one’s

hearing ability can be dramatically overestimated on account of the tendency to produce well-formed responses even if they did not match the perception. Evidence for this includes the higher performance scores for high-context target words despite no appreciable difference in acoustic quality between those words and the same words without context. Given the opportunity, listeners can exploit context to hide true errors in auditory perception. Second, the quantification of intelligibility alone overlooks differences in the speed and efficiency of processing; given the current popularity of prediction-driven frameworks of cognition and linguistic processing, the current results show that signal quality (not just signal content) plays a vital role in language prediction and resulting benefit.

Communication consists of more than simply receiving auditory input; it is also the processing of that input with regard to the expectations of the flow of conversation and the formulation of relevant and creative responses. The lingering effects of prolonged cognitive processing of incoming speech could potentially interfere with some of those *beneath-the-surface* abilities. Specifically, the processing observed after the stimuli for CI listeners and NH listeners with vocoded speech is occurring during the time when it would be reasonable to expect (a) another sentence in the stream of conversation or (b) the preparation of a verbal response more quickly than the 2-s delayed response elicited in this study. With these situations in mind, the translational goal of this study is to highlight a new potential area for auditory outcome measures that extend beyond intelligibility. If a new speech-processing strategy or noise-reduction algorithm elicits shorter latency of effort release for a large group of patients (as quantified here or in some other novel way), it would establish that such a treatment supports faster or more efficient language processing and could prepare patients to succeed better in natural conversation. Future work can explore these issues more directly by varying the type of response elicited or by introducing various types of competing noise/speech that could interfere with poststimulus language processing.

To conclude, the results of the current study suggest that predictive processing does not merely facilitate faster and more accurate responses; it also reduces the *effort* required to understand speech, and the speed of this process is driven at least partly by signal quality or distortion. In light of the rapidity of speech, hearing impairment or some other challenge (e.g., listening in noise, listening to a nonnative speaker, and listening in a reverberant room) can impede a person’s ability to quickly predict upcoming speech by exploiting context cues, and could thus have cascading implications for the effort need for everyday communication, and for thorough audiological assessment.

Author Notes

Franzo Law II, Stefanie Kuchinsky, and Ruth Litovsky contributed valuable ideas to this project.

Acknowledgments

Two anonymous reviewers are thanked for their insightful and helpful feedback on an earlier draft.

Declaration of Conflicting Interests

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

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Institutes of Health–National Institute on Deafness and Other Communication Disorders (NIH-NIDCD R03 DC014309 to Matthew Winn; NIH-NICHHD R01 DC003083 to Ruth Litovsky), by a core grant to the Waisman Center (NIH-NICHHD P30 HD03352), and the University of Wisconsin–Madison Department of Surgery. The author is also supported by the NIH division of Loan Repayment.

References

- 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.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. doi:10.1016/j.tics.2007.05.005.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *Lme4: Linear mixed-effects models using Eigen and S4*. R Package Version 1.1–7. Retrieved from <http://CRAN.R-project.org/package=lme4>.
- 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.
- 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*(March), 32–48. doi:10.1121/1.2017541.
- Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., . . . Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology*, *18*(1), 36–47. doi:10.1159/000343189.
- Boeüx, C., de Balthasar, C., Kós, M.-I., & Pelizzone, M. (2003). Electrical field interactions in different cochlear implant systems. *The Journal of the Acoustical Society of America*, *114*(4), 2049–2057. doi:10.1121/1.1610451.
- Chow, W. Y. (2013). *The temporal dimension of linguistic prediction* (Doctoral Dissertation). University of Maryland College Park.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Science*, *36*, 1–73. doi:10.1017/S0140525X12000477.
- Danermark, B., & Gellerstedt, L. C. (2004). Psychosocial work environment, hearing impairment and health. *International Journal of Audiology*, *43*, 383–389.
- Federmeier, K. D., McLennan, D. B., & Ochoa, E. D. E. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*, 133–146. doi:10.1017.S004857720139203X.
- Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*(2), 1150–1163. doi:10.1121/1.1381538.
- Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, *6*(1), 19–27. doi:10.1007/s10162-004-5024-3.
- Gow, D. W. Jr (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163–179. doi:10.1037//0096-1523.28.1.163.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several years later. *Journal of the Acoustical Society of America*, *87*, 2592–2605. doi:10.1121/1.399052.
- Grimby, A., & Ringdahl, A. (2000). Does having a job improve the quality of life among post-lingually deafened Swedish adults with severe-profound hearing impairment? *British Journal of Audiology*, *34*, 187–195.
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568. doi:10.1016/j.wocn.2010.08.002.
- Hess, E. H. (1972). Pupillometrics. In N. S. Greenfield, & R. A. Sternbach (Eds.), *Handbook of psychophysiology* (pp. 491–531). New York, NY: Holt, Rinehart & Winston.
- Hicks, C., & Tharpe, A. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech and Hearing Research*, *45*, 573–584. doi:10.1044/1092-4388(2002/046).
- Hornsby, B. (2013). The effects of hearing aid use on listening effort and mental fatigue association with sustained speech processing demands. *Ear and Hearing*, *34*, 523–534. doi:10.1097/AUD.0b013e31828003d8.
- Järvelin, M. R., Mäki-Torkko, E., Sorri, M. J., & Rantakallio, P. T. (1997). Effect of hearing impairment on educational outcomes and employment up to the age of 25 years in

- northern Finland. *British Journal of Audiology*, 31, 165–175. doi:10.3109/03005364000000019.
- Kamil, R., Genter, D., & Lin, F. (2015). Factors associated with the accuracy of subjective assessments of hearing impairment. *Ear and Hearing*, 36, 164–167. doi:10.1097/AUD.000000doi0000000075.
- 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.
- 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.
- Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7(1), 18–27. doi:10.1177/1745691611427305.
- Lau, E. F., & Nguyen, E. (2015). The role of temporal predictability in semantic expectation: An MEG investigation. *Cortex*, 68, 8–19. doi:10.1016/j.cortex.2015.02.022.
- Lau, E., Gramfort, A., Hmlinen, M., Kuperberg, G. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *The Journal of Neuroscience*, 33(43), 17174–17181.
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98(1), 74–88. doi:10.1016/j.bandl.2006.02.003.
- Lazard, D. S., Vincent, C., Venail, F., van de Heyning, P., Truy, E., Sterkers, O., . . . Blamey, P. J. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PLoS One*, 7(11), 1–11. doi:10.1371/journal.pone.0048739.
- Lowenstein, O., & Loewenfeld, I. (1964). The sleep-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences*, 117, 142–156. doi:10.1111/j.1749-6632.1964.tb48169.x.
- Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24, 279–284. doi:10.1177/0963721415570732.
- Mahr, T., Mcmillan, B. T. M., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350. doi:10.1016/j.cognition.2015.05.009.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during Word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- Mattys, S., Davis, M., Bradlow, A., & Scott, S. (2012). Speech recognition in adverse conditions: A review. *Language & Cognitive Processes*, 27, 953–978.
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. doi:10.1016/0010-0285(86)90015-0.
- McClelland, J., Mirman, D., Bolger, D., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, 38, 1139–1189. doi:10.1111/cogs.1214.
- 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. *Quarterly Journal of Experimental Psychology*, 58, 22–33.
- McGarrigle, R., Munro, K., Dawes, P., Stewart, A., Moore, Barry, J. G. & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group “white paper”. *International Journal of Audiology*, 53(7), 433–440. doi:10.3109/14992027.2014.890296.
- McGinley, M., David, S., & McCormick, D. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, 187, 179–192.
- McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, 131(1), 509. Retrieved from <http://doi.org/10.1121/1.3664087>.
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. New York, NY: CRC Press.
- Moberly, A. C., Lowenstein, J. H., & Nitttrouer, S. (2016). Word recognition variability with cochlear implants: “Perceptual attention “versus “ auditory sensitivity,” 37, 14–26.
- Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M. & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an Internet-based national survey on hearing. *International Journal of Audiology*, 48(10), 684–691. doi:10.1080/14992020902962421.
- Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech Language and Hearing Research*, 56, 1075–1084. doi:10.1044/1092-4388(2012/12-0074).
- 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.
- Qiuyuan, J., Richer, F., Wagoner, B., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology*, 89, 506–540. doi:10.1111/j.1469-8986.1985.tb01645.x.
- Rabbit, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica*, 476(Suppl): 167–176.
- 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(31doi: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, S99–S105. doi:10.1080/14992020802301167.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge

- during speech perception. *Journal of Neuroscience*, 32(25), 8443–8453. doi:10.1523/JNEUROSCI.5069-11.2012.
- ten Bosch, L., Oostdijk, N., de Ruiter, J. (2004). Turn-taking in social dialogues: temporal, formal and functional aspects. In *Proceedings of the ninth conference on speech and computer (SPECOM)* (pp. 454–461). St. Petersburg, Russia.
- Tavano, A., & Scharinger, M. (2015). Prediction in speech and language processing. *Cortex*, 68, 1–7. doi:10.1016/j.cortex.2015.05.001.
- Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, 24(3), 761–766. doi:10.1037/a0014802.
- Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, 8, 577. doi:10.3389/fnhum.2014.00577.
- Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, 52(1), 23–36. doi:10.1016/j.ijpsycho.2003.12.03.
- Wagner, A., Toffanin, P., & Başkent, D. (2015). How hard can it be to ignore the pan in pan da? Effort of lexical competition as measured in pupil dilation. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow.
- Wendt, D., Dau, T., & Hjortkjaer, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*, 7, 345. doi:10.3389/fpsyg.2016.00345.
- Wilhelm, B., Wilhelm, H., & Lüdtke, H. (1999). Pupillography: Principles and applications in basic and clinical research. In M. Kuhlmann, & J. Böttcher (Eds.), *Pupillography: Principles, methods and applications* (pp. 1–10). München, Germany: Zuckschwerdt Verlag.
- Wingfield, A., McCoy, S., Peelle, J., Tun, P., & Cox, L.C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology*, 17, 487–497.
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, 131(2), 1465. doi:10.1121/1.3672705.
- 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.000000000000145.
- Winn, M. B., Rhone, A. E., Chatterjee, M., & Idsardi, W. J. (2013). The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in Psychology; Auditory Cognitive Neuroscience*, 4(4), 1–13. doi:10.3389/fpsyg.2013.00824.
- 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.
- Zekveld, A., & Kramer, S. (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., Rudner, M., Kramer, S., Lyzenga, J., & Rönnerberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, 8, 88. doi:10.3389/fnins.2014.00088.