

Slower Speaking Rate Reduces Listening Effort Among Listeners With Cochlear Implants

Matthew B. Winn, and Katherine H. Teece

Objectives: Slowed speaking rate was examined for its effects on speech intelligibility, its interaction with the benefit of contextual cues, and the impact of these factors on listening effort in adults with cochlear implants.

Design: Participants ($n = 21$ cochlear implant users) heard high- and low-context sentences that were played at the original speaking rate, as well as a slowed ($1.4\times$ duration) speaking rate, using uniform pitch-synchronous time warping. In addition to intelligibility measures, changes in pupil dilation were measured as a time-varying index of processing load or listening effort. Slope of pupil size recovery to baseline after the sentence was used as an index of resolution of perceptual ambiguity.

Results: Speech intelligibility was better for high-context compared to low-context sentences and slightly better for slower compared to original-rate speech. Speech rate did not affect magnitude and latency of peak pupil dilation relative to sentence offset. However, baseline pupil size recovered more substantially for slower-rate sentences, suggesting easier processing in the moment after the sentence was over. The effect of slowing speech rate was comparable to changing a sentence from low context to high context. The effect of context on pupil dilation was not observed until after the sentence was over, and one of two analyses suggested that context had greater beneficial effects on listening effort when the speaking rate was slower. These patterns maintained even at perfect sentence intelligibility, suggesting that correct speech repetition does not guarantee efficient or effortless processing. With slower speaking rates, there was less variability in pupil dilation slopes following the sentence, implying mitigation of some of the difficulties shown by individual listeners who would otherwise demonstrate prolonged effort after a sentence is heard.

Conclusions: Slowed speaking rate provides release from listening effort when hearing an utterance, particularly relieving effort that would have lingered after a sentence is over. Context arguably provides even more release from listening effort when speaking rate is slower. The pattern of prolonged pupil dilation for faster speech is consistent with increased need to mentally correct errors, although that exact interpretation cannot be verified with intelligibility data alone or with pupil data alone. A pattern of needing to dwell on a sentence to disambiguate misperceptions likely contributes to difficulty in running conversation where there are few opportunities to pause and resolve recently heard utterances.

Key words: Cochlear implants, Listening effort, Speech perception, Speech rate, Time expanded speech.

(*Ear & Hearing* 2021;42;584–595)

Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, USA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

Copyright © 2020 The Authors. *Ear & Hearing* is published on behalf of the American Auditory Society, by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

INTRODUCTION

One of the hallmark pieces of advice when speaking with a person with hearing impairment is “don’t speak louder, speak more slowly.” Previous research supports this, as speech intelligibility improves for individuals with hearing loss when speaking rate is slower (Gordon-Salant & Fitzgibbons 1993; Schneider et al. 2005; Lessa & Costa 2013). In addition to intelligibility, another crucial aspect of listening with hearing loss is listening effort. People with hearing loss report more effort when listening (McCoy et al. 2005; Alhanbali et al. 2017; Hughes et al. 2018), and problems relating to effort are thought to be connected to other consequences such as increased prevalence of mental fatigue (Bess & Hornsby 2014), need for recovery after work (Nachtegaal et al. 2009), and withdrawal from social situations (Hughes et al. 2018). Listening effort is a multidimensional construct (Francis et al. 2016; Pichora-Fuller et al. 2016; Alhanbali et al. 2019) that is likely too complex to reveal direct connections between specific laboratory tasks and long-term experiences. Still, studies aimed at unpacking the multiple components of effort—particularly in speech perception—can potentially help to explain the difficulties experienced by people with hearing impairment in ways that might not be readily accessible in tests where the primary outcome measure is repetition accuracy (i.e., intelligibility) scores. In the current study, slower speech rate was hypothesized to not only yield the aforementioned benefits of higher intelligibility scores but also reduce listening effort and increase the benefit of contextual cues.

A sizeable number of the participants in the aforementioned study by Hughes et al (2018) wore cochlear implants (CIs). Although CIs have been a highly successful treatment for those with hearing loss, they remain limited by degraded sound quality, particularly in the frequency domain. As a result, CI listeners are quite variable in their ability to recognize speech, with some struggling with very poor intelligibility scores (Holden et al. 2013, 2016). CI listeners also show elevated and prolonged listening effort compared to listeners with normal hearing (NH), as well as diminished release from effort when sentences have semantic coherence (Winn 2016). Following up on that finding, listeners with CIs are the focus of the current study, where the use of contextual cues is further examined as it is affected by speaking rate. It should be noted, however, that the issues of sentence perception, speaking rate, and contextual cues likely cut across many kinds of hearing loss, including those who wear hearing aids and those who do not use any devices.

Through various studies cutting across multiple subfields in cognitive psychology, effort has been shown to be a dynamic construct that is best measured over time (Bradshaw 1968; Cavanagh et al. 2014; Vogelzang et al. 2016; McCloy et al 2017; Francis et al. 2018; Kadem et al 2020). Explicit time-series design of listening effort tasks therefore could provide extra insight that might not be revealed via peak or summarized effort

values alone. Time-series measurements of listening effort offer information that is complementary to intelligibility scores. For example, variations in pupil dilation have been linked to specific events during sentence repetition tasks and have been hypothesized to correspond to ongoing uncertainty and a process of language ambiguity resolution (Vogelzang et al. 2016; Winn 2016; Winn & Moore 2018). There are also pupillary signatures of solving mathematical problems or other learning tasks (Bradshaw 1968; Cavanagh et al. 2014). Time-series analysis is also a potentially foundational aspect of describing speech perception as an incremental and rapid process of decomposition (Tanenhaus et al. 1995), and a task in which the brain should be thought of as an active predictor rather than a passive receptor (Wild et al. 2012).

Slow Speech and “Clear” Speech

An entire subfield of literature is focused on the benefits of speaking more clearly (Smiljanić & Bradlow 2009), with experimental results showing better intelligibility of words (Ferguson 2012), sentences (Gilbert et al. 2014), and longer passages (Smiljanić & Bradlow 2008) when talkers are encouraged to speak clearly. Clear speech aids not only word recognition but also memory encoding for older adults (Smiljanić & Chandrasekaran 2013), non-native speakers of a language (Keerstock & Smiljanic 2019; Borghini & Hazan 2020), and normal-hearing adults hearing speech in noise (Van Engen et al. 2012; Gilbert et al. 2014). These results suggest that “stimuli that are easier to process will also be remembered better” (Gilbert et al. 2014). Furthermore, Van Engen et al. (2012) suggested that the presence of semantic coherence in the speech further enhances the benefit of clear speech for recognition memory, a result later corroborated and extended by Keerstock and Smiljanic (2019), who tested listeners hearing clear speech in a non-native language. Although clear-speech benefits are repeatedly shown in the literature, questions remain about exactly why slower speaking rate is beneficial and if it impacts effort as well as intelligibility. A study by Müller et al. (2019) using listeners with NH found that faster speech elicited larger pupillary responses, suggesting greater effort in addition to poorer intelligibility. In that study, syntactic complexity did not produce a substantial change in effort; the current study examines speech-rate changes along with semantic context rather than syntactic complexity and focuses on listeners with hearing loss. Borghini and Hazan (2020) have measured changes in pupil dilation in NH listeners resulting from changes in clear speaking style crossed with sentence plausibility as well as whether the listener was a native speaker of the target language. They found strong effects of native language and speaking style, but surprisingly no effects of sentence plausibility, which they described as a specific case of semantic context (with plausible sentences such as “The talented artist drew a picture” and implausible/anomalous sentences such as “The vegetables open a difficult hat”). In the current study, we approach semantic context differently, with all sentences being plausible, but only some sentences having predictable words (described further in the Methods section).

There is no single acoustic property that is the defining feature of clear speech (Sommer et al. 2019), but one of the most consistent acoustic properties is a reduced speaking rate. This property is of particular interest in the current study for two

reasons: (1) the common anecdotal advice given by audiologists that speaking slowly is more important than speaking loudly, and (2) temporal dimensions of speech are thought to be particularly important for listeners with hearing impairment who use CIs (Shannon 2002) or those who hear a degraded signal generally (Shannon et al. 1995). In the current study, we take a simplified approach to slower speech by applying a uniform time warping, leaving pauses, nonuniform time expansion, and other kinds of prosody manipulations for later study (to be elaborated in the Discussion).

There is very little literature examining clear speech benefits among people who use CIs. Liu et al. (2004) measured a 37.8 percentage-point (or 4.2 dB signal to noise ratio benefit) clear speech advantage for sentence perception in noise among better-performing CI recipients, which was disproportionately higher than that for listeners with NH. In the current study, we expand on the results of Liu et al. (2004) by specifically focusing on slower speaking rate (rather than clear speech overall) and also by measuring corresponding changes in listening effort resulting from that change in rate.

The Use of Semantic Context

Semantic context in speech perception is a focus of the current study because it is thought to be more essential for people with hearing impairment compared to their peers with typical healthy hearing. In the current study, we take the approach of using all grammatically correct sentences that either contain or do not contain internal predictability or semantic coherence (as done previously by Bilger et al. 1984; Pichora-Fuller et al. 1995; Schneider et al. 2005) rather than use coherent versus anomalous sentences (c.f., Stine & Wingfield 1987; Borghini & Hazan 2020). The goal of this choice is to maintain the listeners’ expectation that what they hear should make sense and should be processed as normal language.

Previous studies have shown that context improves intelligibility scores among people with hearing loss (Pichora-Fuller et al. 1995; Holmes et al. 2018), including those who use CIs (Winn 2016) or those who are listening to spectrally degraded speech (Patro & Mendel 2016). However, CI recipients appear to require more time to use context than NH listeners do and demonstrate less release from effort when there is context (Winn 2016; Winn & Moore 2018). Furthermore, the benefit of context to reduce effort is fragile for CI listeners; it shrinks or completely disappears when the moment after a sentence is disrupted by noise or another utterance (Winn & Moore 2018). The specific aspect of how context affects effort *over time* is a focal point in the current study, which exploits time-series pupillary measures to track effort in the moments during and after a sentence.

Summary and Hypotheses

The questions in the current study are whether the ability of CI listeners to use context to increase intelligibility is mediated by the speaking rate of the stimulus, whether speaking rate affects listening effort overall, and whether the benefit of context to reduce effort is mediated by speaking rate. The study used a 2 × 2 design where there was slower and faster speech and high-context or low-context sentences in each rate. There were four main hypotheses in this study: (1) On the basis of numerous previous reports, we hypothesized that the slower speaking rate would promote better intelligibility scores for listeners

who use CIs. (2) Because previous studies showed reduced listening effort among signals that are more intelligible (Zekveld et al. 2010; Koelewijn et al. 2012; Winn et al. 2015) and Müller et al. (2019) showed lower effort for slower speech among NH listeners, we expected reduced listening effort for CI listeners for slowly spoken sentences compared to faster sentences. (3) On the basis of previous observations of reduced benefit of semantic context among CI listeners (Winn 2016; Winn & Moore 2018), we hypothesized that slower speaking rate would more clearly lead to reduction of listening effort resulting from context, because the contextual information would be more intelligible. (4) Although NH listeners show reduced listening effort before a high-context sentence is complete (Winn 2016), CI listeners were hypothesized to show the benefit after the sentence was complete.

METHODS AND MATERIALS

Participants

Data were collected in 21 adults with CIs (age range, 23–82 years; average, 61 years). Two were excluded from data analysis because of poor camera tracking or excessive data loss. Demographic information for the included participants is listed in Table 1. All participants were native speakers of North American English. All participants were able to converse freely during face to face communication, and none reported cognitive or language-learning difficulties. All but one participant acquired hearing loss after language acquisition; the sole perilingually deafened individual has very good speech intelligibility and was deemed capable of performing well enough to be included in the group. The median length of CI use was 6 years, with a range of 1 to 28 years. Of the participants whose data were used,

12 were bilaterally implanted and 7 were unilaterally implanted. Two participants routinely wore a hearing aid in the ear contralateral to unilateral implantation. All were tested using their everyday listening settings, except that the participants with hearing aids were asked to remove the aids during testing; one of these participants preferred to use her hearing aid during testing and was permitted to do so.

Stimuli

Stimuli included a subset of the Revised Speech Perception in Noise (R-SPiN) materials (Bilger et al. 1984) used previously by Winn and Moore (2018). Each stimulus is a grammatically correct English sentence that contains a sentence-final word that is either predictable or not predictable based on the earlier-occurring words. The subset of sentences was selected to contain clear examples of high-context sentences (e.g., “Let’s decide by tossing a coin”; “He wiped the sink with a sponge”) that contain current colloquial language and which did not involve emotional or evocative language, since that would likely influence pupil dilation in a way that reflects something other than listening effort. The low-context sentences (e.g., “He wants to talk about the risk”; “We could consider the feast”) were randomly selected from the entire set of low-context sentences from the R-SPiN corpus. In this study, there were a total of 114 high-context and 118 low-context sentences, with an average of one more low-context sentence per block.

Speech Rate Changes

Speech rate was systematically controlled via the pitch-synchronous overlap-add algorithm implemented in the Praat software (Boersma & Weenink 2018). This technique involves dividing the speech into successive chunks corresponding to pitch periods and replicates or deletes those chunks with overlap. This process maintains spectral envelope and allows control over duration and pitch. For the stimuli in the current study, only duration was manipulated; pitch contours were kept unchanged. There were two speech rates tested: the original rate (“Original”) and a version where the final duration was 140% of the original (“Slow”). The original stimuli were also processed through the pitch-synchronous overlap-add algorithm to ensure that all stimuli were sent through the same processing pipeline.

Procedure

Participants sat in a sound-treated room 50 cm from a single loudspeaker. They viewed a monitor that displayed a simple gray background with a cross in the middle. The luminance of the screen was set at a dark gray (40% of the linear distance between black and white) to avoid large pupil dilations from accommodating a black screen and to avoid eye irritation from viewing a bright screen. During each trial, a warning beep alerted the onset of an upcoming stimulus. Two seconds later, the stimulus began. The cross on the screen remained red throughout the trial until it turned green 2 sec following the offset of each stimulus. This color change served as the prompt for participants to give their verbal response. The instructions were simply to listen to the sentence, and then repeat the whole sentence at the color prompt, giving a best guess when uncertain. Following the end of the participant’s verbal response, the experimenter scored the response and waited until the pupil size returned to baseline before initiating the next trial. This waiting time was typically 5 to 8 sec. The

TABLE 1. Demographics of CI participants

| Listener | Sex | Age | Device Type | Implanted Ear(s) | Etiology of Deafness | Years CI Exp. |
|----------|-----|-----|------------------|------------------|----------------------|---------------|
| C118 | F | 30 | Cochlear | Bilateral | Idiopathic | 7.5 |
| C119 | F | 23 | Cochlear | Bilateral | Idiopathic | 17.5 |
| C121 | M | 52 | Cochlear | Right | Idiopathic | 23 |
| C126 | F | 72 | Med-EI | Bilateral | Idiopathic | 5.5 |
| C127 | F | 73 | Cochlear | Right | Genetic | 7 |
| C130 | M | 66 | Med-EI | Right | Genetic | 1 |
| C131 | F | 70 | Cochlear | Right | Mixed HL | 5.5 |
| C132 | M | 81 | Cochlear | Right | Otosclerosis | 4.5 |
| C134 | F | 63 | Cochlear | Bilateral | Idiopathic | 6 |
| C136 | M | 82 | Advanced Bionics | Left | Genetic | 3.5 |
| C137 | F | 59 | Cochlear | Bilateral | Mixed HL | 2.5 |
| C138 | F | 60 | Advanced Bionics | Bilateral | Idiopathic | 28 |
| C139 | F | 61 | Advanced Bionics | Bilateral | Genetic | 7.5 |
| C140 | F | 46 | Cochlear | Bilateral | Genetic | 2.5 |
| C141 | F | 73 | Advanced Bionics | Right | Genetic | 7 |
| C142 | F | 74 | Cochlear | Bilateral | Idiopathic | 5 |
| C143 | F | 64 | Cochlear | Bilateral | Infection | 3 |
| C144 | F | 62 | Cochlear | Bilateral | Measles | 16 |
| C145 | M | 54 | Cochlear | Bilateral | Meniere’s Disease | 6 |

CI, cochlear implant; F, female; HL, hearing loss; M, male.

time interval between successive trial onsets was roughly 18 sec. Stimulus presentation was conducted with custom MatLab software, which interfaced with an SR Research Eyelink 1000 plus eye tracker sampling at 1000 Hz.

Each testing session began with a set of five sentences to familiarize the listener with the pace of the test and the style of sentences that they would hear. Following the practice, there were four blocks of 29 sentences each, alternating speech rate between blocks. After 15 trials, the screen informed the participant that they could take a break; usually participants preferred to simply continue testing. Within a test block, the speech rate was consistently either original or slow, but the ordering of high- and low-context sentences was pseudo-randomized such that no more than three of the same sentence type were presented consecutively. The order of Original–Slower–Original–Slower speech rate was counterbalanced with Slower–Original–Slower–Original across the pool of participants. The experiment took between 45 and 60 min, depending on each participant's pace and need for breaks.

Analysis

Intelligibility • Because of the special nature of the R-SPiN sentences having a carrier phrase and a target word, intelligibility scoring followed the procedure used previously by Winn (2016), where errors on the final target word were tallied in their own category, and a separate tally was kept of errors among any of the words leading up to that final target word. This contrasts with a style of scoring individual key words, as it was determined that there was not a clear criterion for determining key words or to determine equality of key word value across sentence types.

Intelligibility performance for the lead-up words in each trial was quantified as 1 (all correct) or 0 (at least one error) and was estimated using a generalized mixed-effects binomial logistic model that estimated the log-odds of achieving a correct score as a function of various factors. The fixed effects included speech rate, context, and the interaction between rate and context. Each of these fixed effects were also interacted with the presence or absence of an error on the target word in the trial. Each of these fixed effects and fixed-effect interactions was also declared as a random effect within each listener.

Performance for the target words was modeled using the same set of factors described above, except that the interacting factor of intelligibility was performance on the lead-up words (i.e., binary scoring of the lead-up words was used as a factor in the model that accounted for performance on the target word).

Pupillometry • Consistent with arguments laid out by Winn and Moore (2018), data were examined by way of slope of pupil dilation relative to the end of the utterance. Following the description in their paper, the data were first cleaned using a multistep process that first involved identifying short stretches of missing pupil-size data attributable to blinks, expanding those stretches and then linearly interpolating over them. The data were then low-pass filtered at 5 Hz and decimated to maintain one sample every 40 ms. Baseline pupil diameter was calculated during the 1 sec preceding sentence onset for each trial, and all data points within each trial were transformed to represent proportional (divisive) change relative to that baseline. Trials with excessive (>40%) missing data or contaminated baselines were dropped from the dataset using an automated procedure that identified red-flag patterns such as ± 3 SD differences in mean pupil dilation or baseline dilation or significant sloping values

during baseline. Number of dropped trials was individually variable (i.e., some participants had cleaner eye tracking data), but generally the number of dropped trials was 6 to 9 and was not found to be associated with speech rate or context.

Following the analysis procedures used by Winn (2016) and Winn and Moore (2018), pupil data were divided into two windows of analysis corresponding to the “listening” portion and the “wait” portion (i.e., “retention interval”). The window 1 had a variable duration depending on the speaking rate, as pupils tend to dilate about 0.7 sec after the onset of an auditory stimulus. The window 1 for the original-rate speech began at -1.8 sec relative to stimulus offset and ended 0.7 sec relative to sentence offset. The window 1 for the slow-rate speech began at -2.4 sec relative to stimulus offset and ended 0.7 sec relative to sentence offset. Analysis window 2 was the same regardless of rate, as the stimulus would have been completed, and the subsequent repetition task was equivalent across speaking rates. Window 2 began at 0.7 sec relative to stimulus offset and continued to 2.4 sec relative to stimulus offset, when it was determined that the pupils showed a signature of a nonauditory response to the color change in the visual prompt. Slope of change in pupil dilation was obtained by including time as a predictor in the statistical model. Maximally defined mixed-effects models were used to account for dependence of individuals' slopes across speaking rates and listening conditions, thus shrinking the estimated variance and increasing the power to robustly and fairly detect differences across these conditions while minimizing Type I error rates.

Analysis of the Effect of Speaking Rate Overall • The effect of speaking rate overall was not statistically compared during window 1, as the nonequivalence in stimulus time would not enable fair comparisons. However, pupil dilation within window 2 was compared across rates, using fixed effects of time (slope), rate, and the interaction between time and rate. Each of these terms were also included as subject-level and item-level random effects. Statistical analysis was conducted using the R software (version 4.0.0; R Core team 2020), using the lme4 (version 1.1.23; Bates et al. 2015) and lmerTest (version 3.1.2; Kuznetsova et al. 2017) packages. The prevailing model took the following form using R notation:

```
lmerTest :: lmer(pupil ~ Time * Rate + # Main effects
  (1 + Time * Rate | Listener) + # Random listener effects
  (1 + Time * Rate | Stimulus), # Random item effects
  data = window_2)
```

Analysis of Context During Window 1 (Listening) • The effect of context within window 1 was estimated separately for each speaking rate, to foster fair comparisons among stimuli with the same general duration. For each set of data (Original and Slow), the prevailing model took the following form (written in R notation):

```
lmerTest :: lmer(pupil ~ Time * Context + # Main effects
  (1 + Time * Context | Listener) + # Random effects
  (1 + Time | Stimulus), # Random effects
  data = window_1_slow_rate) # Data subset (slow
  or original rate)
```

Analysis of Context During Window 2 (the Waiting Time After the Sentence) • For window 2, it was possible to compare all possible combinations of context and speaking rate in a unified model, as the timing landmarks in the study were all equivalent following the end of the utterance. There were fully crossed fixed effects of time (slope), rate, context, and all interactions within these factors. The same set of factors were used as subject-level random effects, along with additional random effects of intercept, time (slope), rate, and the interaction between time and rate per item. There was no random effect of context per item as context was an inherent property of each stimulus. The prevailing model took the following form using R notation:

```
lmerTest :: lmer(pupil ~ Time
                + Rate + Context + # Main effects
                Time : Rate + Time : Context
                + Rate : Context + # twos - way interactions
                Time : Rate : Context + # three - way interaction
                (1 + Time + Rate + Context + # Random effects
                Time : Rate + Time : Context + # Random interactions
                Rate : Context | Listener) + # per Listener
                (1 + Time * Rate | Stimulus), # Random item effects
                data = window_2) # Data subset
```

RESULTS

Intelligibility

Intelligibility scores are displayed in Figure 1, split by sentence component, speaking rate, and context type. The lead-up portion (i.e., the words before the sentence-final target word) of low-context sentences was repeated with roughly 87% accuracy, regardless of speaking rate. The lead-up portion of high-context sentences was also repeated back with 87% accuracy for slow-rate sentences; this performance dropped to 81% for original-rate sentences. There was substantially better performance for

sentence-final target words when they were preceded by coherent semantic context, with a roughly 30 percentage-point increase for high-context words observed in both the original and slower-rate sentences.

Table 2 displays results of the generalized linear mixed-effects model used to describe the intelligibility results. For lead-up words, there were no statistical main effects of speaking rate (β_2 ; $p = 0.718$) nor context (β_3 ; $p = 0.768$). However, there was a decrease in performance for lead-up words when there was an error on the target word (β_4 ; $p = 0.025$), which was an even stronger effect when the stimulus was a high-context sentence (β_7 ; $p < 0.001$). For sentence-final target words, the main effect of context was strong (β_{11} ; $p < 0.001$). The lack of clear statistical effect of speech rate on target words did not change when the stimuli were high-context sentences (β_{13} ; $p = 0.232$).

Just as for the analysis of lead-up words, the target word performance in slower-rate stimuli was reduced when there was an error elsewhere in the sentence, for both the low-context (β_{12} ; $p = 0.006$; estimate -0.86) and even more for the high-context stimuli (β_{15} ; $p < 0.001$; $\beta_{12} + \beta_{15}$ interaction estimate of -0.86 to $2.639 = -3.5$), where the beta estimate was over four times as large (-3.5 versus -0.86 log odds). Neither of these interaction effects of error on lead-up were statistically different for original-rate speech (β_{14} ; β_{16}).

Pupillometry

Main Effect of Speaking Rate • Figure 2 shows change in pupil dilation over time for both the original- and slow-rate sentences, averaged over both low- and high-context sentence types. The pattern suggests that peak magnitude and peak latency (relative to sentence offset) were not different across the speaking rates. Differences in the rising onset slopes in the data are reflective of the overall difference in stimulus duration, consistent with multiple previous studies (Winn & Moore 2018; Müller et al. 2019; Borghini & Hazan 2020).

Table 3 shows the output of the mixed-effects model accounting for differences in pupil size during window 2, which was

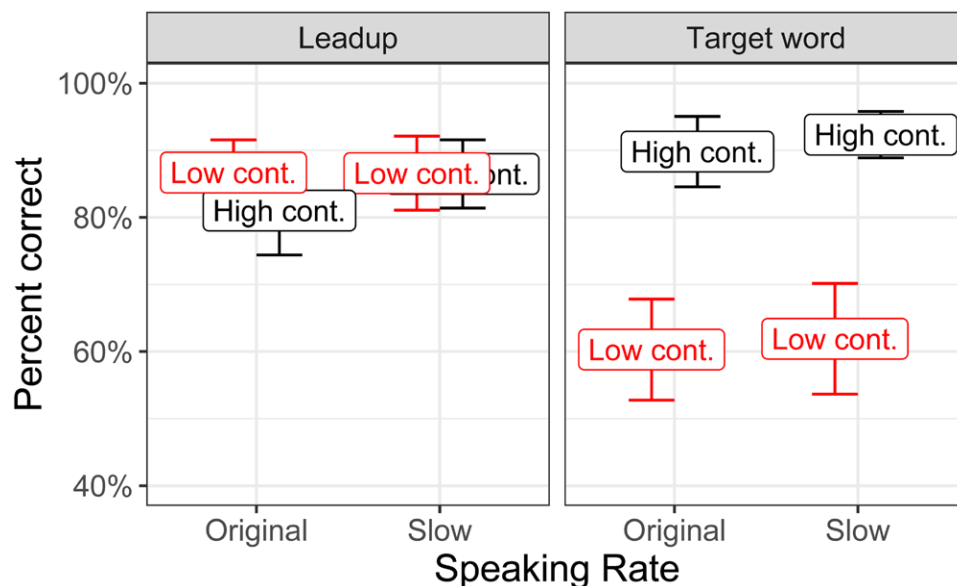


Fig. 1. Intelligibility scores. Panels for scores for lead-up (left) and target (right) words in sentences heard at original or slowed rates. High-context and low-context performance is represented in black and red, respectively. Error bars represent ± 2 SEM.

TABLE 2. Results of generalized (binomial logistic) linear mixed-effects models accounting for intelligibility of lead-up words ($\beta 1$ – $\beta 8$) and target words ($\beta 9$ – $\beta 16$)

| | | Estimate | st. err. | z | p |
|------------------------------------|---|----------|----------|-------|--------|
| Lead-up words | | | | | |
| $\beta 1$ | (Intercept) | 2.809 | 0.422 | 6.66 | <0.001 |
| $\beta 2$ | Rate (original) | –0.154 | 0.426 | –0.36 | 0.718 |
| $\beta 3$ | Context (high) | –0.116 | 0.392 | –0.30 | 0.768 |
| $\beta 4$ | Error-on-target | –0.919 | 0.410 | –2.24 | 0.025 |
| $\beta 5$ | Rate (original): Context (high) | –0.120 | 0.519 | –0.23 | 0.817 |
| $\beta 6$ | Rate (original): error-on-target | 0.238 | 0.549 | 0.43 | 0.664 |
| $\beta 7$ | Context (high): error-on-target | –2.697 | 0.706 | –3.82 | <0.001 |
| $\beta 8$ | Rate (original): Context (high): error-on-target | –1.331 | 1.240 | –1.07 | 0.283 |
| Sentence-final target words | | | | | |
| $\beta 9$ | (Intercept) | 0.731 | 0.206 | 3.55 | <0.000 |
| $\beta 10$ | Rate (original) | –0.246 | 0.146 | –1.69 | 0.091 |
| $\beta 11$ | Context (high) | 3.302 | 0.404 | 8.18 | <0.001 |
| $\beta 12$ | Error-on-lead-up | –0.860 | 0.314 | –2.74 | 0.006 |
| $\beta 13$ | Rate (original): Context (high) | 1.123 | 0.940 | 1.20 | 0.232 |
| $\beta 14$ | Rate (original): error-on-lead-up | 0.432 | 0.449 | 0.96 | 0.336 |
| $\beta 15$ | Context (high): error-on-lead-up | –2.639 | 0.604 | –4.37 | <0.001 |
| $\beta 16$ | Rate (original): Context (high): error-on-carrier | –1.243 | 1.119 | –1.11 | 0.267 |

st. err. is SEM estimation.

0.7 to 2.4 sec relative to stimulus offset, corresponding to the data in Figure 2. This model accounted only for overall pupil dilation without regard to sentence context. There was no detectable effect of time in the original-rate sentences, implying a flat slope ($\beta 2, p = 0.429$). Slower speech rate was associated with a steeper negative offset slope compared to the model default ($\beta 4, p = 0.041$).

Effects of Context

Figure 3 illustrates changes in pupil dilation over time split by speaking rate (panels) and context type (color within each panel). During window 1 (listening period), effects of context were modeled separately for original-rate and slow-rate speech because the stimuli occupied different amounts of time; Table 4 contains the details of these models, which used low-context

stimuli as the default condition and used time as a model term to reflect the slope of change over time. For original-rate speech, there was no main effect of context ($\beta 3$) on the main intercept term ($\beta 1$). The slope for low-context sentences was statistically greater than zero ($t = 7.72; p < 0.001; \beta 2$), reflecting ongoing increases in pupil dilation during window 1. However, the interaction between time and context was not statistically detectable ($\beta 4, p = 0.57$), showing that the slope of pupil dilation during the window 1 was not affected by context, replicating previous results with CI listeners and maintaining contrast with previous results in NH listeners (Winn 2016). The same pattern of effects (main effect of slope, but no effect of context on the intercept or the slope terms) was also observed in the model for the slow rate as well (Table 4, $\beta 5$ through $\beta 8$). These results did not support the hypothesis that slower speaking rate in these stimuli would facilitate “online” benefit from context to reduce effort during the listening process.

Table 5 shows details of the mixed-effects model that accounted for effects of sentence context over time on pupil dilation during time window 2, which was 0.7 to 2.4 sec relative to stimulus offset. The default configuration of the model corresponds to high-context slow-rate speech; the intercept ($\beta 1$) was significantly greater than zero, simply reflecting the presence of pupil dilation at the peak just after the offset of the sentence.

TABLE 3. Linear mixed-effects model accounting for change in pupil size in window 2 (time 0.7 to 2.4 sec relative to stimulus offset)

| | Term | Estimate | st. err. | df | t | p |
|-----------|--------------------------------|----------|----------|-------|-------|--------|
| $\beta 1$ | Intercept— (original rate) | 0.109 | 0.012 | 19.75 | 9.03 | <0.001 |
| $\beta 2$ | Time (slope, original rate) | –0.004 | 0.005 | 19.33 | –0.81 | 0.429 |
| $\beta 3$ | Rate (slow) | –0.001 | 0.011 | 24.19 | –0.11 | 0.912 |
| $\beta 4$ | Time: rate (slow) | –0.012 | 0.006 | 20.44 | –2.19 | 0.041 |

st. err. is SEM estimation; df is degrees of freedom estimated using the Satterthwaite approximation (implementation by Kuznetsova et al. 2017).

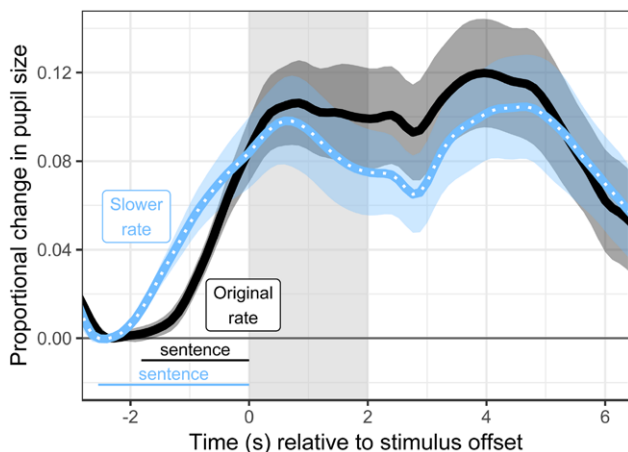


Fig. 2. Proportional change in pupil dilation over time for sentences spoken at the original rate (black line) or a slower rate (blue line with dots). Width of the error ribbon represents ± 2.1 SEM. The vertical gray shaded region represents the silent interval between stimulus offset and the visual prompt for listeners to repeat the sentence.

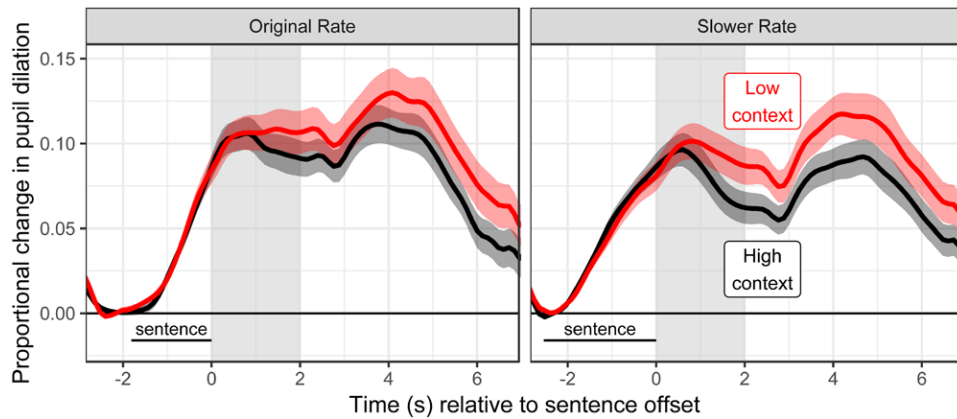


Fig. 3. Proportional change in pupil dilation over time. Panels illustrate data for sentences spoken at the original rate (left) or at a slower rate (right). Low-context sentences are displayed with red lines, and high-context sentences are displayed with black lines. Width of the error ribbon represents ± 2.1 SEM. The gray shaded region represents the silent interval between stimulus offset and the visual prompt for listeners to repeat the sentence.

The intercept was not affected by speech rate (β_3) or context (β_4), nor the interaction between the two (β_7).

Pupil size shrank back toward baseline following high-context slow-rate sentences, as the slope (“time”) term was statistically less than zero (Table 5, β_2). Changing the speech rate from slower to original resulted in a shallower slope (β_5 , indicating prolonged listening effort) and removing context also produced a shallower slope (β_6). Interestingly, changing speech rate produced an effect that was equivalent in magnitude to the effect produced by semantic context, although the variability in the speech-rate effect was enough to make it statistically weaker (β_5 ; $p = 0.07$) than the context effect (β_6 ; $p = 0.004$). The three-way interaction between time, speech rate, and context was not statistically detectable (β_8), indicating that the benefit of context to steepen the downward slope of pupil dilation was statistically the same for slow-rate speech as it was for original-rate speech.

None of the effects reported here were statistically different when analyzing only trials in which the sentences were repeated correctly. Full models with intelligibility interactions are available in Supplemental Digital Contents 1 <http://links.lww.com/EANDH/A717> (overall dilation), 2 <http://links.lww.com/EANDH/A718> (window 1), and 3 <http://links.lww.com/EANDH/A719> (window 2).

com/EANDH/A718 (window 1), and 3 <http://links.lww.com/EANDH/A719> (window 2).

Sentence Postprocessing

We are particularly interested in modeling changes in pupil dilation after the peak—reflected in the offset slope during window 2—because it likely reflects ongoing uncertainty in processing the utterance. Engelhardt et al. (2010) measured the slope of changes in pupil dilation following a specific word that disambiguated pronouns, and Bradshaw (1968) measured reductions in pupil size locked to the time of solving mental arithmetic. In the current study, there were no such specific word landmarks, but there were clear differences in slope following the general landmark of sentence offset. Figure 4 visualizes the transformation of these offset slope data into the summarized modeled values that were listed in Table 5. There is a sequence of points corresponding to the actual linear slopes for each listener (as X 's), the transformation of those slope values when incorporating random effects for listeners and items (as open points), converging on the group estimated slope values (larger filled points in the center of each panel) accounting for combined random effects of items and listeners, including

TABLE 4. Linear mixed-effects model accounting for change in pupil size for sentences during window 1 (time -2.2 to 0.7 sec relative to stimulus offset for slow rate, and between -1.8 and 0.7 sec for original rate)

| | Estimate | st. err. | df | t | p |
|-----------------------------------|----------|----------|-------|-------|--------|
| Original rate | | | | | |
| β_1 Intercept (Low context) | 0.084 | 0.010 | 26.90 | 8.07 | <0.001 |
| β_2 Time (slope) | 0.053 | 0.007 | 24.51 | 7.72 | <0.001 |
| β_3 High-context | 0.002 | 0.009 | 62.62 | 0.27 | 0.785 |
| β_4 Time: high-context | 0.003 | 0.005 | 55.10 | 0.59 | 0.557 |
| Slow rate | | | | | |
| β_5 Intercept (low context) | 0.086 | 0.012 | 23.94 | 7.16 | <0.001 |
| β_6 Time | 0.038 | 0.006 | 22.27 | 6.74 | <0.001 |
| β_7 High-context | -0.005 | 0.011 | 62.61 | -0.48 | 0.631 |
| β_8 Time: high-context | -0.003 | 0.005 | 49.35 | -0.57 | 0.570 |

st. err. is SEM estimation; df is degrees of freedom estimated using the Satterthwaite approximation (implementation by Kuznetsova et al. 2017).

TABLE 5. Linear mixed-effects model accounting for change in pupil size in window 2 (time 0.7 to 2.4 sec relative to stimulus offset)

| | Estimate | st. err. | df | t | p |
|---|----------|----------|-------|-------|--------|
| β_1 Intercept (slow-rate, high-context) | 0.108 | 0.017 | 20.79 | 6.46 | <0.001 |
| β_2 Time (slope) | -0.022 | 0.004 | 21.51 | -4.91 | <0.001 |
| β_3 Original-rate | 0.003 | 0.014 | 26.20 | 0.22 | 0.825 |
| β_4 Low-context | 0.000 | 0.010 | 52.15 | 0.03 | 0.980 |
| β_5 Time (slope): original-rate | 0.012 | 0.006 | 21.16 | 1.91 | 0.070 |
| β_6 Time (slope): low-context | 0.012 | 0.004 | 29.85 | 3.11 | 0.004 |
| β_7 Original-rate: low-context | -0.003 | 0.012 | 56.42 | -0.24 | 0.814 |
| β_8 Time: original-rate: low-context | -0.001 | 0.005 | 46.65 | -0.19 | 0.846 |

st. err. is SEM estimation; df is degrees of freedom estimated using the Satterthwaite approximation (implementation by Kuznetsova et al. 2017).

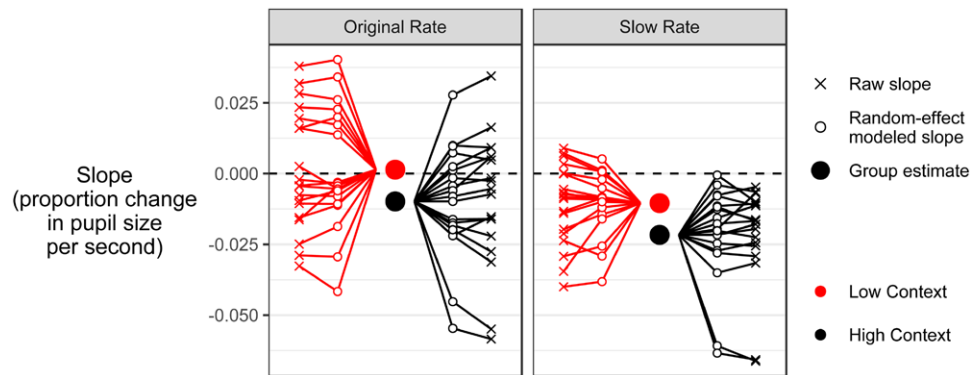


Fig. 4. Slope of pupil dilation following post-stimulus peak, corresponding to “window 2” from the analysis. Raw slope values are indicated by X’s, which transition to the open points, which reflect the values adjusted by random-effect structure accounting for dependence of data within speaking rate and context (and their interaction) per listener and also accounting for random effects of items in the stimulus set. The large filled points in the center reflect the estimated group-level slopes, which are the mean of the random-effects estimated slopes.

random-effect interactions. Points falling below the zero line indicate a negative slope, which in this case would be a sign of success, as it would reflect recovery back toward resting-state pupil size.

Figure 4 shows greater variability in pupil dilation slopes following the original-rate speech compared to slow-rate speech. CI listeners were more similar in their processing of slower speech than in their processing of faster speech (SDs of 0.0109 and 0.0218, respectively, for slower and original-rate speech). Seven listeners (one third of the group) showed markedly higher slopes for original-rate speech that all fell toward flat/negative slope values when speech was slowed. This distribution of data suggests that the ability or inability to handle faster speaking rate is specific to the individual rather than a universal feature of using a CI. The benefits of slower speech could be described as bringing the entire group into a similar range by mitigating the difficulty of the one-third of participants who struggled most with the original-rate speech.

Context-Related Effort Release

An additional analysis was conducted that calculated “effort release,” quantified as the proportional reduction of pupil dilation for high-context sentences relative to low-context sentences. Consistent with previous studies that analyzed the effect of context on pupillary measures of listening effort by our research group (Winn 2016; Winn & Moore 2018), effort release was quantified as the linear difference between low-context and high-context pupil responses, divided by the peak pupil dilation in the low-context condition. The advantage of this approach is that every proportional change is expressed with reference to the individual’s peak pupil dilation in the task, thus self-normalizing for individual differences in pupil reactivity. Additionally, since the prompt-related short-term deflection in pupil dilation around 2.4 sec is time-locked and should be equal across all conditions, the calculated difference between two conditions should neutralize it, thus allowing a longer time window without undesirable task-irrelevant deflections in the data. The disadvantage of this approach is that it requires aggregated data to directly compare high-context responses to low-context responses (rather than estimating the outcome for each context type separately) and therefore does not include trial-level data or a random effect of stimulus. Figure 5 shows this

calculation for the slow and original-rate speech, as well as the corresponding measure for listeners with NH, whose data come from the study by Winn (2016).

Statistical modeling of effort release used a time analysis window between 0.7 and 3.3 sec relative to stimulus offset. The reason for this extended offset time was that average verbal reaction time by CI listeners in a similar experiment (where responses were audio recorded) was measured to be 0.6 sec following the response prompt. The offset landmark of 3.3 sec relative to the stimulus offset was determined by taking that 0.6 sec timepoint, adding a customary 0.7 sec to account for the latency of cognitive task-evoked pupil dilation, and accounting for the 2-sec silent retention interval.

Because the morphology of the proportional effort release data was not suitable for a simple linear analysis, effort release was modeled used a third-order orthogonal polynomial mixed-effects model (c.f., Mirman 2014; Winn et al. 2015) so that the linear, quadratic, and cubic changes over time could be

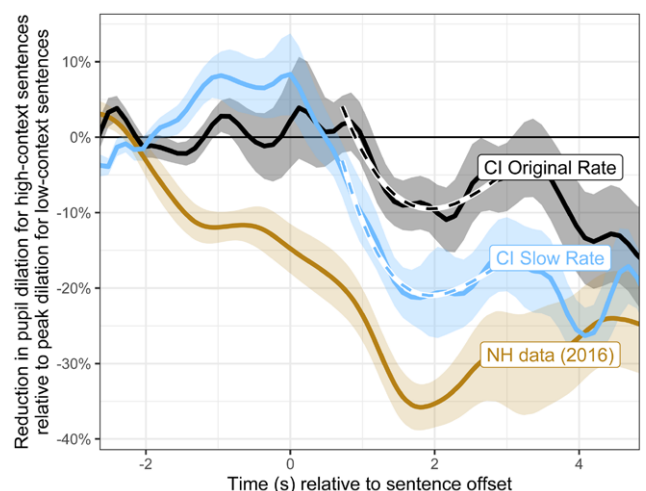


Fig. 5. Difference between pupil dilation responses for low-context and high-context stimuli, divided by the peak dilation in low-context stimuli, representing release from effort related to sentence context. Data in blue and black are split by speaking rate for listeners with cochlear implant (CI), with data for listeners with normal hearing (NH) reproduced from the study by Winn (2016). Dashed lines represent data during the time window that was statistically modeled (Table VI).

estimated independently from each other. The other fixed effects were speech rate and the interactions of speech rate with each of the three polynomial expressions of time. There was maximal random-effects structure (with a random effect declared for each of the fixed effects), in order to guard against inflated risk of Type I errors. The model results revealed no interacting effects of speech rate with any of the time polynomials, thus suggesting a simpler model without those two-way interactions. A second model was constructed with simple fixed effects of speech rate (as an intercept term) and the three time polynomials, and it was found to be a more parsimonious model according to a likelihood-ratio (χ^2) test using the Akaike Information Criterion (Akaike 1974).

The model took the following form, expressed using R notation:

```
lmerTest :: lmer(release ~ poly1 +
  poly2 + poly3 + Rate +                               #Main effects
  (1 + poly1 + poly2 + poly3 + Rate | Listener),         #Random effects
  data = window_2)                                     #Data subset
```

Table 6 shows the summarized output for both the full and reduced (parsimonious) models of effort release, but we discuss the reduced model only. The intercept for slow-rate speech was lower than that for original-rate speech, implying more benefit from context when speech rate was slower. In the mixed-effects model, this pattern was arguably statistically detectable (Table 6, β_{13} : $t = 2$; $p = 0.06$). When excluding the random effects, the statistical effect was larger ($t = 8.5$; $p < 0.001$), validating the notion that random-effects structure provided a more conservative estimate of effects. The strength of the quadratic and cubic terms reflects the major nonlinearity during the window of analysis (the quadratic term), which is asymmetrical and approaches a second inflection (necessitating the cubic term).

It should be noted that although each model contains test statistics in a unified framework (i.e., each row in the table is part of one model, rather than being an individual statistical test), the presence of multiple models invites caution when deciding to reject a null hypothesis with a borderline test statistic. Although we do not advocate for the stance that test statistics should be

treated in a categorical *all-or-none* fashion, we wish to highlight the presence of multiple statistical models and hence multiple opportunities to identify effects.

DISCUSSION

Main Hypotheses

Slower speaking rate appears to reduce listening effort among CI listeners during the period just after a sentence, to a degree that approximates the effort release obtained by having semantic context in the sentence (Fig. 3, confirming hypothesis #2). There was some evidence that slower speaking rate increases the benefit of context as measured by release from effort (which would validate hypothesis #3), although this evidence emerged for only one of the two approaches to the analysis. Compared to data from NH listeners who participated in a similar study (Winn 2016), the CI listeners in this study showed context-related release from effort later in time, confirming hypothesis #4. Slowing the speaking rate did not appear to make this effort release substantially earlier, which is consistent with a framework of CI speech perception operating with a chronic disposition of delaying commitment to a perception until after an utterance is over (c.f., Farris-Trimble et al. 2014; McMurray et al. 2017), heavily loading importance onto the extra moment after a sentence.

Surprisingly, there were no major effects of intelligibility that were associated solely with slower speaking rate, thus not confirming hypothesis #1, and thus showing inconsistency with previous literature. Despite the lack of major changes in intelligibility scores, perhaps the benefit of slower speech is a reduction in the need to continue processing the previous utterance, indicated by the greater recovery back toward baseline just after the offset of slower sentences (Figs. 2, 3). Additionally, the effects of intelligibility could be masked by the opportunity for listeners to retroactively “repair” the original-rate utterances by guessing at a sensible word so that they can report a well-formed answer despite not hearing it clearly. This “extra moment” after a sentence has previously been identified as a fragile moment during speech perception by CI users (Winn & Moore 2018), as disturbance of auditory attention just after a sentence can

TABLE 6. Linear mixed-effects model accounting for reduction in pupil size for high-context sentences relative to low-context sentences, as a proportion of peak dilation in low-context sentences

| | Term | Estimate | st. err. | df | t | p |
|--|---------------------------------|----------|----------|-------|-------|-------|
| β_1 | Intercept—(original rate) | 0.058 | 0.034 | 18.26 | 1.73 | 0.101 |
| β_2 | Time (slope, linear) | 0.058 | 0.084 | 17.89 | 0.69 | 0.497 |
| β_3 | Time (quadratic) | -0.154 | 0.052 | 17.79 | -2.96 | 0.009 |
| β_4 | Time (cubic) | 0.033 | 0.037 | 17.93 | 0.89 | 0.388 |
| β_5 | Rate (slow) | 0.109 | 0.055 | 18.13 | 2.00 | 0.060 |
| β_6 | Time (slope, linear): slow rate | 0.052 | 0.123 | 17.70 | 0.42 | 0.676 |
| β_7 | Time (quadratic): slow rate | -0.023 | 0.065 | 17.83 | -0.36 | 0.725 |
| β_8 | Time (cubic): slow rate | 0.023 | 0.052 | 17.87 | 0.43 | 0.670 |
| More-parsimonious model without rate:polynomial interactions | | | | | | |
| β_9 | Intercept—(original rate) | 0.058 | 0.034 | 18.00 | 1.72 | 0.102 |
| β_{10} | Time (slope, linear) | 0.084 | 0.056 | 18.00 | 1.51 | 0.150 |
| β_{11} | Time (quadratic) | -0.166 | 0.043 | 18.00 | -3.89 | 0.001 |
| β_{12} | Time (cubic) | 0.044 | 0.022 | 22.59 | 2.00 | 0.057 |
| β_{13} | Rate (slow) | 0.109 | 0.055 | 18.00 | 2.00 | 0.061 |

The window of analysis was time 0.7 to 3.3s relative to stimulus offset. st. err. is SEM estimation; df is degrees of freedom estimated using the Satterthwaite approximation (implementation by Kuznetsova et al. 2017).

disrupt processing of the sentence. Therefore, finding of effort during the moment after a sentence in the current study lends further support to the notion that everyday continuous speech could be more challenging than what is estimated from single-sentence stimuli, because opportunities to continue processing the previous utterance are rare or costly when the next sentence begins right away.

A Closer Look at Intelligibility

Analysis of intelligibility in the current study revealed that errors in early parts of the sentence are not independent from errors on later parts of the sentence. As such, the “repair” process mentioned earlier could be just as likely detrimental as it is beneficial. Previous research by Marrufo-Pérez et al. (2019) has shown that target words are repeated with systematically lower accuracy when preceding contextual words are misperceived. As opposed to an ideal situation where later words were perceived more accurately because of a buildup of related contextual words preceding them, Marrufo-Pérez et al. found that later words were perceived less accurately, specifically because of inaccuracies in perceiving the earlier words. Although this result seems intuitive in retrospect, it demonstrates that the presence of semantic context should be considered beneficial only in situations where contextual words are perceived correctly, unlike the noise-masked conditions used by Marrufo-Pérez et al. or the situation in the current study which used listeners with CIs.

If a sentence begins before the listener has completed processing the previous one, there could be difficulties that do not emerge in intelligibility scores when testing only one utterance at a time. Without a time-series physiological measure such as pupillometry, EEG, etc., or a behavioral method that is sensitive to auditory processing after a target utterance (Capach et al. 2019), this phenomenon of delayed language processing is likely not detectable using conventional approaches (e.g., single utterances). Other studies using eye-tracking paradigms with CI listeners have corroborated the finding of delayed language processing, but at the lexical level (Farris-Trimble et al. 2014; McMurray et al. 2017). Such experiments hold value for bridging the gap between “normal” outcome measures for individuals who struggle in real-life listening situations where the stream of speech lacks sufficient silent gaps to reprocess recent words.

Speaking Rate and Context

Super-additive effects of speech rate and context on pupil dilation were not detected in the full statistical model, despite aggregated data showing more release from effort obtained from sentence context when the speech was slower (Fig. 5; Table 6). There are considerations and trade-offs to each style of statistical modeling, such as the directness of estimating an effect derived from comparisons of aggregated data versus the trial-level data that contains extra statistical power but which also demands additional model complexity. As for previous studies by our research team (Winn 2016; Winn & Moore, 2018), disambiguation and resolution of language processing was analyzed using the relative reduction of pupil size for high-context compared to low-context sentences, which was a measure that was internally normalized by each participant’s peak pupil reactivity during the task. This measure has now shown to land within a stable range across three studies, with replicable differences between listeners with NH and CI.

Müller et al. (2019) found that slower speaking rate resulted in smaller *peak* pupil dilations, while the current study found differences after the peak but not at the peak. For speech manipulation, Müller et al. used the same algorithm as the current study, but had other potentially important methodological differences. First, they used a $\pm 25\%$ (lengthening/shortening) duration manipulation, whereas we used the original durations and a 40% lengthening. Second, they used the Oldenburg Linguistically and Audiologically Controlled Sentences (Uslar et al. 2013), which have a rigid sentence structure (commonly referred to as “matrix” sentences) as opposed to the somewhat more syntactically diverse R-SPiN sentences in the current study. Perhaps, most importantly, the listeners in the study by Müller et al. had NH, whereas the listeners in the current study used CIs. Borghini and Hazan (2020) measured context-related differences in peak pupil dilations in NH listeners across clear and conversational speech, whereas CI listeners have been found to show little to no differences in peak pupil dilation resulting from context (Winn 2016; Winn & Moore 2018) or speech rate (the current study). Collectively, these studies suggest that there could be important qualitative differences between listening to speech with a hearing loss versus listening to non-native speech and also differences between listening to genuine clear speech versus artificially slowed speech.

Figure 4 gives the impression that for low-context original-rate speech, there are two groups of listeners—one group with positive slopes and another group with flat or negative slopes, following sentence offset. This appearance of bimodality was not verified statistically, perhaps because it consisted of only 7 and 14 listeners, respectively. Each of the 7 participants with highest slopes in the original-rate low-context sentences (Fig. 4, left panel, red points) showed a substantially reduced slope in the slow-rate condition (Fig. 4, right panel, red points), and the group standard deviation in slopes was reduced by roughly 50% (0.021 to 0.011) when comparing original rate to slow rate, implying a partial neutralization of some of the individual differences that extended into the upper (undesirable) range of slopes.

Reflecting on Clear Speech and Slowed Speech

True “clear speech” is likely to provide even more substantial benefits than those measured in the current study, because it would involve more than simple uniform time expansion. For example, there are phoneme-level changes such as hyperarticulated final consonants (Picheny et al. 1985) and vowels (Picheny et al. 1985; Ferguson & Kewley-Port 2002; Smiljanić & Bradlow 2009) including greater spectral dynamics in vowels (Lam et al. 2012; Ferguson & Quené 2014). However, expansion of vowel space alone is not sufficient to support intelligibility. For example, McCloy (2013) found that it was not absolute vowel space, but rather the difference between prosodically stressed and prosodically unstressed vowel space that promoted better intelligibility. Unnatural emphasis of unstressed syllables is therefore potentially detrimental. This nuance is not always observed in common analyses of vowel space that only account for the hyperarticulated edges of the vowel space. McCloy’s analysis further suggests that acoustic differentiation of vowel segments might promote better access to prosodic emphasis. Further to this point, clear speech also tends to involve a wider dynamic pitch range and more prosodic phrasing (Smiljanić & Bradlow 2008), reinforcing the idea that clearer speech involves stronger cues for emphasis within an utterance (de Jong 1995).

Despite the differences between real clear speech and the time-expanded speech used in this study, there is a large range of potential applications of time-expanded speech. Sometimes it is not feasible for a talker to change the speaking style to be clearer, but it is possible to artificially slow down the rate of previously recorded speech to potentially provide benefit to listeners with CIs, listeners with other kinds of hearing loss, or non-native speakers of a language. For example, it is common to encounter video-recorded class lectures, educational videos for children, workplace safety videos, flight safety videos, employment orientation materials, and other materials related to employment and educational equity. Time-expansion algorithms currently in use for internet video streaming and podcast players might potentially be of great value to those who face challenges in comprehending occupational or educational media.

Apart from time expansion, the rate of speech information can be slowed by the insertion of pauses within an utterance. However, those pauses are beneficial only when they are inserted at syntactically appropriate places (Wingfield et al. 1999). Considering the fragility of the extra moment after a sentence for listeners with CIs (Winn & Moore 2018; Gianakas & Winn 2019; Capach et al. 2019), these pauses possibly represent an opportunity for the listener to resolve perceptual ambiguities before the speech resumes, which might protect against unsustainable buildup of several threads of ambiguous speech streams. In agreement with this hypothesis, Van Engen et al (2012) showed that more-clearly spoken sentences were not only recalled with greater accuracy, but also that there were fewer false alarms in recognizing previously heard utterances. In other words, speech clarity protected against the likelihood that listeners entertained multiple alternative perceptions that would later be erroneously labeled as actually being heard. Lingering effects of cognitive processing are unlikely to be revealed in short-latency single-utterance scoring in speech perception tests but could play a vital role in speech processing by individuals with hearing loss.

CONCLUSIONS

For listeners with CIs, slower speaking rate appears to reduce the effort that would otherwise continue past the end of a sentence. This result suggests that slower speaking rate results in reduced uncertainty after a sentence, potentially enabling a listener to be more prepared to hear another sentence after the previous one has ended. Testing for this type of prolonged uncertainty likely demands time-series analysis or the use of multiple utterances within a single trial. Having context in a sentence is arguably even more beneficial when the speech rate is slower (supported by one of two separate analyses). Slower speech resulted in overall reduction as well as reduction of individual variability in pupil slope—a measure of cognitive resolution—following the sentence. There are numerous situations where speech could be artificially slowed to potentially provide benefit to listeners with CIs or listeners with other kinds of hearing loss or non-native speakers of a language.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant NIH NIDCD R01 DC017114 (Winn). Data collection was assisted by Emily Hugo, Paula Rodriguez, Hannah Matthys, and Lindsay Williams. The University of

Minnesota stands on *Miní Sóta Makhóche*, the homelands of the Dakhóta Oyáte.

The authors have no conflicts of interest to disclose.

Address for correspondence: Matthew B. Winn, Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Drive SE, Minneapolis, MN 55455, USA. E-mail: mwinn@umn.edu

Received January 31, 2020; accepted August 12, 2020; published online ahead of print October 30, 2020.

REFERENCES

- Alhanbali, S., Dawes, P., Lloyd, S., Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear*, 38, e39–e48.
- Alhanbali, S., Dawes, P., Millman, R. E., Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear*, 40, 1084–1097.
- Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J Statistical Software*, 67, 1–48.
- Bess, F. H., & Hornsby, B. W. (2014). Commentary: Listening can be exhausting—fatigue in children and adults with hearing loss. *Ear Hear*, 35, 592–599.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res*, 27, 32–48.
- Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer [Computer program]. Version 6.0.37*. Retrieved on March 7, 2018. <http://www.praat.org/>
- Borghini, G., & Hazan, V. (2020). Effects of acoustic and semantic cues on listening effort during native and non-native speech perception. *J Acoust Soc Am*, 147, 3783.
- Bradshaw, J. L. (1968). Pupil size and problem solving. *Q J Exp Psychol*, 20, 116–122.
- Capach, N., Neukam, J., Azadpour, M., Sagi, E., Wingfield, A., Svirsky, M. (2019). The “Two-sentence problem”: Communication requires understanding speech blocks longer than a single utterance. Poster presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.
- Cavanagh, J. F., Wiecki, T. V., Kochar, A., Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *J Exp Psychol Gen*, 143, 1476–1488.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J Acoust Soc Am*, 97, 491–504.
- Engelhardt, P. E., Ferreira, F., Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Q J Exp Psychol (Hove)*, 63, 639–645.
- 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. *J Exper Psych: Hum Perc Perf*, 40, 308–327.
- Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *J Speech Lang Hear Res*, 55, 779–790.
- Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, 112, 259–271.
- Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *J Acoust Soc Am*, 135, 3570–3584.
- 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. *Front Psychol*, 7, 263.
- Francis, A., Tigchelaar, L., Zhang, R., Zekveld, A. (2018). Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech. *J Speech Lang Hear Res*, epub, 1–16.
- Gianakas, S., & Winn, M. (2019). Disruption of the benefit of sentence context in listeners with cochlear implants. Poster presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.
- Gilbert, R. C., Chandrasekaran, B., Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *J Acoust Soc Am*, 135, 389–399.

- Gordon-Salant, S., & Fitzgibbons, P. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res, 40*, 423–431.
- Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear, 34*, 342–360.
- Holden, L. K., Firszt, J. B., Reeder, R. M., Uchanski, R. M., Dwyer, N. Y., Holden, T. A. (2016). Factors affecting outcomes in cochlear implant recipients implanted with a perimodiolar electrode array located in scala tympani. *Otol Neurotol, 37*, 1662–1668.
- Holmes, E., Folkeard, P., Johnsrude, I. S., Scollie, S. (2018). Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *Int J Audiol, 57*, 483–492.
- Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear, 39*, 922–934.
- Kadem, M., Herrmann, B., Rodd, J., Johnsrude, I. (2020). Pupil dilation is sensitive to semantic ambiguity and acoustic degradation. *bioRxiv*. <https://doi.org/10.1101/2020.02.19.955609>
- Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *J Acoust Soc Am, 146*, 4604.
- Koelwijn, 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 Hear, 33*, 291–300.
- Koelwijn, 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. *Hear Res, 323*, 81–90.
- Kuznetsova, A., Brockhoff, P., Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *J Stat Softw, 82*, 1–26.
- Lam, J., Tjaden, K., Wilding, G. (2012). Acoustics of clear speech: effect of instruction. *J Speech Lang Hear Res, 55*, 1807–1821.
- Lessa, A. H., & Costa, M. J. (2013). The impact of speech rate on sentence recognition by elderly individuals. *Braz J Otorhinolaryngol, 79*, 745–752.
- Liu, S., Del Rio, E., Bradlow, A. R., Zeng, F. G. (2004). Clear speech perception in acoustic and electric hearing. *J Acoust Soc Am, 116*(4 Pt 1), 2374–2383.
- Marrufo-Pérez, M., Eustaquio-Martín, A., Lopez-Poveda, E. (2019). Speech predictability can hinder communication in difficult listening conditions. *Cognition, 192*, 103992. <https://doi.org/10.1016/j.cognition.2019.06.004>
- McCloy, D. (2013). Prosody, intelligibility and familiarity in speech perception. *Unpublished doctoral dissertation*, University of Washington.
- McCloy, D. R., Lau, B. K., Larson, E., Pratt, K. A. I., Lee, A. K. C. (2017). Pupillometry shows the effort of auditory attention switching. *J Acoust Soc Am, 141*, 2440.
- 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. *Q J Exp Psychol A, 58*, 22–33.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press.
- 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.
- Müller, J. A., Wendt, D., Kollmeier, B., Debener, S., Brand, T. (2019). Effect of Speech Rate on Neural Tracking of Speech. *Front Psychol, 10*, 449.
- 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. *Int J Audiol, 48*, 684–691.
- Patro, C., & Mendel, L. L. (2016). Role of contextual cues on the perception of spectrally reduced interrupted speech. *J Acoust Soc Am, 140*, 1336.
- Picheny, M. A., Durlach, N. I., Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J Speech Hear Res, 28*, 96–103.
- Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97*, 593–608.
- Pichora-Fuller, M.K., Kramer, S., Eckert, M., Edwards, B., Hornsby, B., Humes, L., Lunner, T., Matthen, M., Mackersie, C., Naylor, G., Phillips, N., Richter, M., Rudner, M., Sommers, M., Tremblay, K., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear, 37*, 5S–27S.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing [computer program] Version 4.0.0*. <https://www.R-project.org>
- Schneider, B. A., Daneman, M., Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychol Aging, 20*, 261–271.
- Shannon, R. V. (2002). The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol, 11*, 124–127.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303–304.
- Smiljanic, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *J Acoust Soc Am, 124*, 3171–3182.
- Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Lang Linguist Compass, 3*, 236–264.
- Smiljanic, R., & Chandrasekaran, B. (2013). Processing speech of varying intelligibility. *Proc Meet Acoust, 19*, 1–4.
- Sommers, M., Spehar, B., Tye-Murray, N., Myerson, J., Hale, S. (2019). Age differences in the effects of speaking rate on auditory, visual, and auditory-visual speech perception. *Ear Hear, 41*, 549–560.
- Stine, E. L., & Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychol Aging, 2*, 272–279.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.
- Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *J Acoust Soc Am, 134*, 3039–3056.
- Van Engen, K. J., Chandrasekaran, B., Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS One, 7*, e43753.
- Vogelzang, M., Hendriks, P., Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Lang Cogn Neurosci, 31*, 876–885.
- Wild, C. J., Davis, M. H., Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage, 60*, 1490–1502.
- Wingfield, A., Tun, P. A., Koh, C. K., Rosen, M. J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychol Aging, 14*, 380–389.
- Winn, M. B., Edwards, J. R., Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear, 36*, e153–e165.
- Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear, 20*, 1–17.
- Winn, M. B., & Moore, A. N. (2018). Pupillometry s. *Trends Hear, 22*, 2331216518808962.
- Zekveld, A. A., Kramer, S. E., Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear, 31*, 480–490.