OPEN

Pupil Dilation Reflects Perceptual Priorities During a Receptive Speech Task

Hannah Mechtenberg,[1] Cristal Giorio,[2] and Emily B. Myers[1,3]

Objectives: The listening demand incurred by speech perception fluctuates in normal conversation. At the acoustic-phonetic level, natural variation in pronunciation acts as speedbumps to accurate lexical selection. Any given utterance may be more or less phonetically ambiguous—a problem that must be resolved by the listener to choose the correct word. This becomes especially apparent when considering two common speech registers—clear and casual—that have characteristically different levels of phonetic ambiguity. Clear speech prioritizes intelligibility through hyperarticulation which results in less ambiguity at the phonetic level, while casual speech tends to have a more collapsed acoustic space. We hypothesized that listeners would invest greater cognitive resources while listening to casual speech to resolve the increased amount of phonetic ambiguity, as compared with clear speech. To this end, we used pupillometry as an online measure of listening effort during perception of clear and casual continuous speech in two background conditions: quiet and noise.

Design: Forty-eight participants performed a probe detection task while listening to spoken, nonsensical sentences (masked and unmasked) while recording pupil size. Pupil size was modeled using growth curve analysis to capture the dynamics of the pupil response as the sentence unfolded.

Results: Pupil size during listening was sensitive to the presence of noise and speech register (clear/casual). Unsurprisingly, listeners had overall larger pupil dilations during speech perception in noise, replicating earlier work. The pupil dilation pattern for clear and casual sentences was considerably more complex. Pupil dilation during clear speech trials was slightly larger than for casual speech, across quiet and noisy backgrounds.

Conclusions: We suggest that listener motivation could explain the larger pupil dilations to clearly spoken speech. We propose that, bounded by the context of this task, listeners devoted more resources to perceiving the speech signal with the greatest acoustic/phonetic fidelity. Further, we unexpectedly found systematic differences in pupil dilation preceding the onset of the spoken sentences. Together, these data demonstrate that the pupillary system is not merely reactive but also adaptive—sensitive to both task structure and listener motivation to maximize accurate perception in a limited resource system.

Key words: Anticipation, Cognitive effort, Continuous listening, Pupillometry, Speaking styles, Speech perception.

[1]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, USA; [2]Department of Psychology, Pennsylvania State University, State College, Pennsylvania, USA; and [3]Department of Speech, Language and Hearing Sciences, University of Connecticut, Storrs, Connecticut, 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).

(Ear & Hearing 2024;45;425–440)

This article has received an OSF badge for Open Data.

INTRODUCTION

Imagine the speaking style of a presenter addressing a room full of academics at a busy conference. Now imagine the speaking style when talking with friends afterward. In the first context, intelligibility is prioritized. The speaker may hyperarticulate their stops and fricatives, expand vowel spaces, and speak slower to maximize their audience's understanding. These constraints are lifted when the speaker sits next to a friend and casually discusses how the talk went. Their vowel space contracts, differences between /d/ and/t/ are less obvious, and their speech rate increases (Smiljanić & Bradlow 2009). In these contexts, does the listener, then, allocate effort differently when perceiving clear versus casual speech? We investigated this question—of listening effort and speech type—using pupillometry. Pupillometry captures subtle, autonomic changes in pupil area in response to a stimulus (task-evoked pupil response) and has been linked to cognitive phenomena of arousal, attention, task demand, and effort (Stanners et al. 1979; Beatty 1982; Unsworth & Robison 2015; Paulus et al. 2020). We aimed to characterize how listener effort is modulated by the intelligibility of the incoming signal when listening to clear versus casual speech embedded in background noise and in quiet. Our goal was to investigate two sources of demand in listening (noise and phonetic distinctiveness) while taking into account behavioral response factors such as task difficulty and individual motivation, which may impact the processing of the overall signal (van der Wel & van Steenbergen 2018).

Speech intelligibility is often under siege, with threats emerging both from the environment (i.e., noise) as well as from the speech signal itself. When listening demands rise, the listener must expend effort, in the form of increased cognitive resources, to meet the cost incurred by the challenge of mapping the acoustic signal to meaning (Heald & Nusbaum 2014). Models of active speech processing, like the Ease of Language Understanding model (Rönnberg et al. 2013), suggest that when intelligibility of the incoming speech signal is low, cognitive resources are used to attain accurate perception. While a fully resolved definition of listening effort remains elusive, we consider listening effort as the "resources or energy actually used by a listener to meet cognitive demands" (Peelle 2018, p. 205). This operationalization purposefully untangles effort from demand, as demand refers to the challenges to accurate perception (Francis & Love 2020), while effort is an individual endeavor to overcome (or

<zdoi; 10.1097/AUD.0000000000001438>

not) the many obstacles to speech perception. In the present study, we challenged listeners with two types of naturally occurring demand—background noise and acoustic/phonetic ambiguity—to better characterize the relationship between listening effort and speech intelligibility using pupillometry.

Background Noise, Speech Intelligibility, and Pupillometry

Listening to speech embedded in background noise is more effortful than listening in silence due to the masking noise obscuring crucial pieces of the acoustic signal (Mattys et al. 2012; Holube et al. 2016; Dryden et al. 2017). The difficulty of the listening task, that is, the demand of the listening context, is thought to increase reliance on cognitive processes of working memory and selective attention (Beatty 1982; Verney et al. 2004; Moresi et al. 2008; Jepma & Nieuwenhuis 2011; Ng et al. 2013) to better filter the signal information from noise to increase speech intelligibility (Pichora-Fuller 2003; Koelewijn et al. 2012; Rönnberg et al. 2013). Applying the Ease of Language Understanding model, these additional cognitive resources are only recruited once demand becomes sufficiently high (Rönnberg et al. 2013). It is possible, then, that speech perception in quiet environments is a more automatic, and negligibly effortful, process. A limitation of behavioral studies of speech-in-noise is the inability to measure online changes in listening effort induced by background noise. van der Wel and van Steenbergen (2018) qualify that mean differences in accuracy on a task are not sufficient to adjudicate between conditions with high demand and instances of high effort. Without a direct measure of effort, a task with a high demand, thus requiring a certain amount of effort to overcome, may be behaviorally indistinguishable from a task with perhaps a lower overall demand but for which there was inadequate effort expended to succeed. In other words, a listener might try very hard at a difficult task and arrive at the same performance level as a listener who is not trying very hard at an easy task.

Fortunately, pupillometry, the dynamic measurement of pupil dilation, is thought to capture fluctuations in listening effort during perception of spoken sentences (for review, see Van Engen & McLaughlin 2018). Pupil dilations are responsive to different speech intelligibility levels, primarily induced by variable masking levels and masker types (Koelewijn et al. 2011, 2012; Zekveld & Kramer 2014). The effect of masking noise on pupil dilation is well-documented: in a recent study cataloging the relationship between peak pupil dilation and signal to noise ratios (SNRs) while listening to sentences, the largest pupil dilations were associated with the least signal clarity (−10 dB) while the smallest dilations accompanied the best signal clarity (+5 dB) (Książek et al. 2021), though others have shown peak dilations at less severe SNRs ranging between −16 and 0 dB (Koelewijn et al. 2012; Zekveld & Kramer 2014; Ohlenforst et al. 2018).

There is increasing interest in examining whether the intelligibility of the acoustic-phonetic signal—and its impact on listening effort—can be measured using pupillometry (Colby & McMurray 2021). Across a range of stimulus types, listeners show larger pupil dilations when confronted with more challenging listening conditions. This pattern has been observed for noise-vocoded speech (Winn et al. 2015) and during perception of non-native accented speech compared with natively accented speech (Porretta & Tucker 2019; McLaughlin & Van Engen 2020). Crucially, differences in dilation can be observed even when listeners are quite accurate. McLaughlin and Van Engen (2020) reported larger pupil dilations while listening to the non-native accented speech than for the native productions, even when limiting analyses to accurately perceived sentences, suggesting that pupil dilation is sensitive to variation in listening effort even when no difference is noted in accuracy data. An open question is whether pupillometry can measure even finer-grained differences between two types of naturally occurring speech: clear and casual registers.

The Speech Signal and Intelligibility

Intelligibility depends not only on the properties of background noise but also on the speech signal itself (Mattys et al. 2012). Many speakers adopt a speaking style known as clear speech (Picheny et al. 1986; Ferguson & Kewley-Port 2002; Krause & Braida 2002; Smiljanić & Bradlow 2005). Clear speech contrasts with a “casual” or “conversational” speaking style. Casual speech places less pressure on articulatory precision during production—leading to collapsed vowel spaces as well as increased speaking rate (Smiljanić & Bradlow 2009), unlike clear speech, which has larger vowel spaces that are associated with higher ratings of intelligibility (Bond & Moore 1994; Bradlow et al. 1996; Ferguson & Kewley-Port 2002; Hazan & Markham 2004). In natural speech, there is substantial overlap in the acoustic cues that characterize a given vowel category (Peterson & Barney 1952; Xie & Myers 2018). Therefore, an expanded vowel space reduces this overlap, and in doing so reduces how much an individual vowel token has to “compete” to be categorized correctly, thus reducing processing demand (Xie & Myers 2018) and, theoretically, also reducing listening effort.

Previous studies have investigated the “clear speech benefit”; describing the phenomenon of improved intelligibility for clear speech compared with casual speech in a variety of noisy backgrounds including quiet, wideband and speech-shaped noise, and multi-talker babble (Picheny et al. 1985; Krause & Braida 2002; Liu et al. 2004; Ferguson 2012) and across SNRs (Payton et al. 1994). The clear speech benefit can range from as low as 4.5% to as high as 30% (Gagné et al. 1995), though the benefit is not found for every talker (Lam & Tjaden 2013). Although the clear speech benefit is biggest when sentences are presented in background noise, it also emerges in silence (Rodman et al. 2020); suggesting a blanket “cost” for perceiving casual speech in most situations. Studies of the clear speech benefit have primarily focused on identifying the acoustic, prosodic, and semantic cues that lead to improved intelligibility (Gagné et al. 1995; Krause & Braida 2004; van der Feest et al. 2019), and have yet to explore differences in listening effort directly.

In this study, we used pupillometry during speech perception to investigate the relationship between demand and listening effort. By crossing two levels of environmental demand, (masked versus unmasked), with two levels of signal demand stemming from acoustic properties (clear versus casual speech), we can systematically investigate how listening effort changes based on amount, and type, of demand. This noise manipulation allows for replication of prior studies of pupillometry during speech-in-noise, but also may amplify the potential benefit of clear speech for the listener, because intelligibility benefits for clear speech are known to be increased under challenging listening conditions (Ferguson 2012, see discussion earlier).

We chose to use semantically meaningless sentences to prevent listeners from relying heavily on top-down information such as semantic prediction (van der Feest et al. 2019). Although these sentences present an unusual listening scenario, they force listeners to attend to the bottom-up acoustic-phonetic details of speech rather than leaning heavily on predictive mechanisms. In this way, we aim to measure pupil changes related to the demands of the acoustic-phonetic signal—isolated from the additional demands or benefits of semantic processing.

We hypothesized that: (1) there will be larger pupil dilations for masked speech perception than for unmasked and (2) perception of casual speech will evoke larger pupil dilations than clear speech. Considering that the largest dissociations in pupil responses occur at moderate SNRs (Zekveld & Kramer 2014; Hopstaken et al. 2015), we set our masking to approximate a 75% intelligibility level, or a +5 dB SNR. Recent work has shown that real-world noise levels range between +2 and +16 dB (Wu et al. 2018), making +5 dB SNR optimal for dynamic pupil dilations and to capture ecological listening conditions. Of interest is whether larger pupil responses indexing the increased processing load of casual speech also emerge in the unmasked condition, as recent behavioral findings show a blanket "cost" for perceiving casual speech with and without background noise (Rodman et al. 2020). We blocked all conditions of interest into mini-blocks of 12 trials each following current guidelines (Winn et al. 2018). It is important to note that mini-blocks allowed us to examine listener adaptation to a particular speech type within blocks. Taken together, this design enabled us to examine not only pupil responses to perceptual challenge, but also the dynamics of this response as listeners adjust to the demands of the input.

MATERIALS AND METHODS

Participants

We recruited 63 participants from the community and our institution's Psychology Participant Pool. Thirteen participants failed to complete the experiment due to software failure or for personal reasons. We further excluded 2 participants for having more than 20% missing eyetracking data (a threshold suggested by Winn et al. 2018), resulting in 48 participants for analyses. The age of the remaining participants ranged from 18 to 22 years (mean = 19; 42 females, 6 males). Of our remaining 48 participants*, 6 were Asian, 2 were Black, and 40 were White. Further, 2 participants were Hispanic or Latino while the remainder were not Hispanic or Latino. All participants reported normal or corrected-to-normal vision, normal hearing, and indicated they were native speakers of North American English. Before the experiment, participants provided written informed consent in accordance with our Institutional Review Board and received either course credit or cash as compensation.

*Sample size was primarily determined from a pilot experiment (not reported) of 30 participants that showed a significant difference between the pupil curves associated with perceiving clear versus casual speech. Forty-eight participants also allowed for an equal number of participants in each counterbalancing condition.

Stimuli

Stimuli consisted of 96 semantically nonsensical, but syntactically well-formed sentences (e.g., trout is straight and also writes brass) obtained from a previous study (Xie & Myers 2018). Each sentence was recorded in two speaking styles—clear and casual—and was evaluated for goodness of style membership by a metric described later. A female native speaker of North American English (E.B.M) produced sentences in each speech register, which were sampled at 44.1 kHz and normalized to 70 dB root mean square amplitude. The clear and casual sets were equated on mean pitch and standard deviation of F_0. Sentence duration ranged from 1372 to 3019 msec, with a small difference in average duration (clear = 2294 msec, casual = 2202 msec, $t(95) = -5.08$, $p < 0.001$). Although sentence duration differed between the two speaking styles, we do not expect a small duration difference (<100 msec) to significantly bias results as divergences in pupil dilations are expected while the sentence unfolds due to the delayed onset of the pupillary response. How stimuli length affects the shape of the pupil curve ultimately is an open question (see McLaughlin & Van Engen 2020 for a discussion on the effects of stimuli length on pupil dilation).

To quantitatively assess each sentence's membership to either clear or casual speaking styles, the degree of phonetic competition was calculated for each stressed vowel and then averaged over the full sentence. Broadly, phonetic competition describes how "close," in acoustic space, a given vowel token is to tokens belonging to other categories. A high value indicates that a vowel token is closely surrounded by tokens from other categories (e.g., an /i/ token surrounded by many /I/ tokens) and must compete to be sorted into the appropriate phonetic category. Conversely, a low value suggests that the vowel token is far from tokens of other categories and the demand for categorization is relatively low. For more details on the calculation of phonetic competition, see Xie and Myers (2018). As depicted in the density plot in Figure 1A, casually spoken sentences had greater overall phonetic competition than clear sentences ($t(392) = 7.18$, $p < 0.001$), capturing the collapsed phonetic space characteristic of casual speech. There were no significant differences in intelligibility (clear = 93.7%, casual = 92.4%, $p = 0.15$), as normed via transcription during the original study these stimuli were created for (Xie & Myers 2018).

Target sentences were presented in two types of background competition—silent or masked with a multi-talker babble. The multi-talker babble, developed by Newman (2005), consisted of nine layered female speakers reading passages from several books. Selection of masker streams were randomized for each target sentence and root-mean-square amplitude normalized to 70 dB. The masker began 3000 msec before the onset of the target sentence and continued for 3000 msec after target sentence offset. For the overlapping portion of the masker stream and the target sentence, we selected a moderately difficult SNR of +5 dB, which approximated participants' 75% signal reception threshold for word intelligibility. Manipulation and construction of all auditory stimuli were completed using Praat (Boersma & Weenink 2019).

Procedure

A desk-mounted EyeLink 1000 Plus (SR Research, Ontario, Canada) recorded the elliptical pupil area and location of the right eye. All equipment, including a chin rest to stabilize

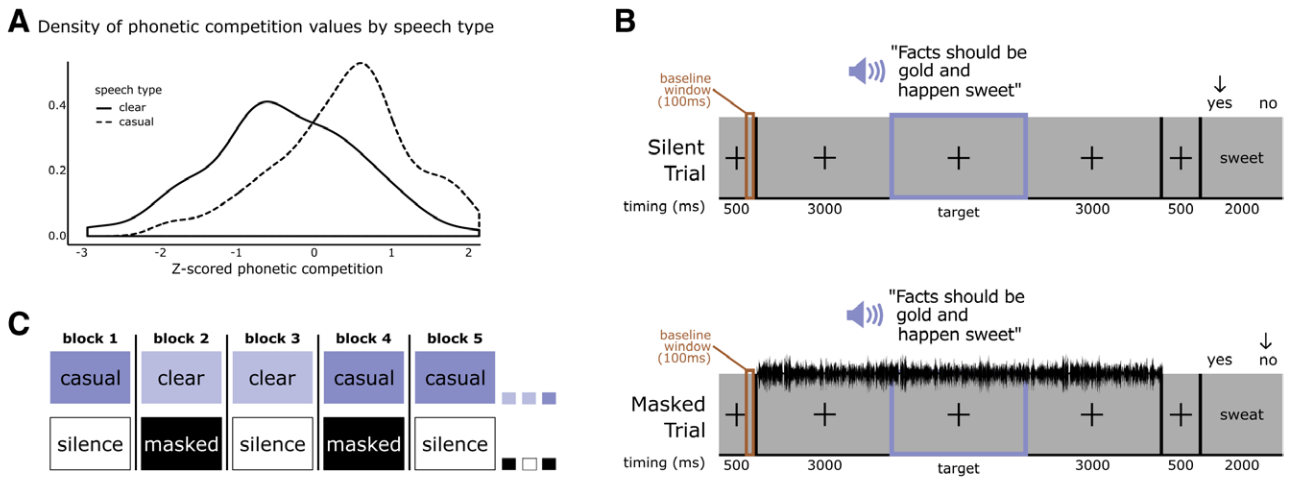

Fig. 1. Methods details. A, Density plot of the *z*-scored phonetic competition values per speech type, with the solid line corresponding to clearly spoken speech and dashed line corresponding to casually spoken speech, highlighting greater phonetic competition in casually spoken sentences compared with clearly spoken sentences. B, Schematics for a silent trial (above) and a masked trial (below). The timing was identical for each trial type, with the only difference being the addition of multi-talker babble 3000 msec before sentence onset which continued until 3000 msec after sentence onset in masked trials. At the conclusion of all trials, participants indicated via key press whether the word on the screen was heard in the presented sentence. C, Schematic depicting the blocked structure of the task. The top row indicates the by-block pattern of casually and clearly spoken speech. The bottom row shows the alternation between silence and masked trials by block. There were 16 blocks total blocks, with 12 trials per block.

participant's heads, was positioned in accordance with recommendations from SR Research. The sampling rate was set to 500 Hz. We began data collection with a nine-point calibration procedure, but then changed to a five-point procedure for the remaining participants after observing similar calibration fidelity. We do not expect differences in data quality between the two calibration procedures as the current experiment was not concerned with gaze direction. Stimulus presentation was controlled by SR Research Experiment Builder (version 2.2.299; SR Research, Ontario, Canada) on a Mac mini (Macintosh, Macmini6,1) computer running Windows 7 with Boot Camp and displayed on a BenQ XL2420-B 24-in monitor. Auditory stimuli were presented binaurally by an external soundcard (M-Track Plus, 24-bit) through over-ear headphones (Sony, MDR-7506, 63 ohm), and volume was consistent across participants. Room luminance was brought to a consistent and moderate level ($M = 78.5$ lumens).

At the start of the session, after obtaining informed consent, participants sat in a sound-attenuated testing room and were administered the QuickSIN; a test widely used to evaluate speech recognition in noise (Killion et al. 2004). The QuickSIN roughly determines each participant's 50% signal reception threshold. During the test, participants listened to five sets of five sentences and repeated them out loud. After each set, the background noise level increased. The accuracy was calculated and incorporated into each participant's threshold. Final thresholds were calculated for each participant (rounded to the nearest 5 dB) to which +5 dB was added to approximate a moderately difficult 75% signal reception threshold (Desjardins & Doherty 2014). We originally intended to individually adjust SNRs by-participant to ensure that the ~75% speech reception threshold was consistent across participants, but because our participants were fairly homogeneous, a +5 dB SNR was used for all participants during masked trials.

The pupillometry experiment consisted of 16 total blocks, each containing 12 trials. Background masker type alternated by block, such that there were eight silent blocks and eight masked blocks. The speaking style of the target sentence, clear or casual, was consistent within a block. After the initial block, speaking style alternated in a two-block pattern (i.e., A [clear], B [casual], B, A, A, B, B, A). Starting in block 9, we flipped the speech type alternation pattern for the second experiment half (i.e., B, A, A, B, B, A, A, B). See Figure 1C for schematic. To note, this experiment design resulted in unintended downstream methodological and empirically interesting effects which we discuss at length in later sections. Participants heard a total of 192 sentences, 96 in clear speech and 96 in casual speech. Each sentence was heard twice, once in each speaking style. All sentences were presented first (in either a clear or casual style) before being repeated in the opposite speech style in the second half of the experiment. The starting masker type and speech type were counterbalanced across participants, as well as which sentences were masked. Thus, if a given sentence was masked in the first half for one speaking style, it was also masked in the second half for the other speaking style. We modeled experiment half (first versus second half) to account for the repetition of sentences across the experiment. Thus, the factor of "experiment half" (see Results) could be interpreted as reflecting sentence repetition, or alternatively, effects related to the passage of time in the experiment (e.g., fatigue).

At the start of each block, we completed calibration and validation procedures to maximize eyetracking data quality, and each trial began with drift correction before recording. Participants took breaks in between each block to reduce fatigue and eye strain. The background color of the screen (Gray; CMYK: 33.2%, 21.6%, 31.4%, 4.3%) remained constant across all calibration, validation, drift correction, and trial stages. All text (fixation cross, probe word) was displayed in black in the center of the screen. A central fixation crossappeared on the screen at all times except during the probe word decision. Before the task began, participants read instructions on the screen.

All trials began with a 500 msec quiet fixation. Then, there was a 3000 msec delay that was either silent or filled with the multi-talker masker. After the delay, the target sentence played. Immediately following the target sentence there was an additional delay of 3000 msec that matched the background condition (i.e., silent or masked) as the initial delay. Before the probe word decision, there was a quiet 500 msec period. At the end of each trial, a probe word appeared on the screen and participants indicated via keyboard press if they heard the visually presented word in the previous sentence or not. Probe words queried the final word of the target sentences. For half of the trials, the probe word was heard in the sentence (e.g., "brass" versus "grass" where "brass" was in the sentence). The trial concluded when participants made a response or else when the response window timed out after 2000 msec. See Figure 1B for comprehensive schematic of both trial types. Upon completion of the experiment, participants were debriefed and appropriately compensated for their time.

Pupil Data Preprocessing

Raw pupil data were preprocessed using the "GazeR" package (Geller et al. 2020) in R (R Core Team 2021). Blinks were automatically identified in GazeR using the "saccades" package (Malsburg 2015). First, we identified blink windows as 100 msec before and 200 msec after missing data (blinks), based on standard blink definitions (Kwon et al. 2013). Then, the pupil trace was smoothed using a five-point moving average filter, and next we applied linear interpolation to the blink windows. Visual inspection of the pupil trace for a single trial confirmed adequate interpolation. For subtractive baselining, we decided on an a priori baseline window beginning 100 msec before target sentence onset. However, after plotting the entire trial baselined to a 100 msec[†] window immediately preceding the 3000 msec delay[‡], we discovered that the pupil traces began to pull apart within our a priori baseline window (see Fig. 3A, inside the red boxes for visual depiction). We will further discuss the methodological implications of baseline window choice in the Discussion. Further, in a secondary post hoc analysis, we analyzed pupil dilations in a 2000 msec window before target sentence onset (which overlapped with our a priori baseline window). The median pupil area within the baseline window, for each trial, was used in a subtractive baselining procedure (Mathôt et al. 2018; Reilly et al. 2019) to correct for tonic pupil drift.

Next, participants and trials were removed from further analysis if missing more than 20% raw pupil data (Winn et al. 2018). To clarify, participants were not removed on the basis of missing more than 20% of total trials, but for missing more than 20% of pupil data across the entire experiment. Two participants were excluded from further analyses as a result. Artifact and outliers were removed using visual histogram inspection and median absolute deviation in accordance with current suggestions (Winn et al. 2018; Geller et al. 2020). Out of 9216 total trials, 529 were removed. Last, we aligned the pupil trace to two different events in preparation for growth curve analysis (GCA) and downsampled to 100 msec bins. The first event, decided a priori, was aligned to the onset of the target sentence. The second, defined post hoc, was a 2000 msec long pre-target window aligned to 2000 msec before the onset of the target sentence (referred to as the pre-target window). The preprocessing code along with the raw data and analyses code may be found here.

Data Analysis

Probe Word Responses • We removed all nonresponses (i.e., timed out) for both accuracy and response time. A total of 531 trials (in addition to the 529 trials removed for missing greater than 20% of pupil data) were removed, leaving 8156 observations. Accuracy[§] and response time data were analyzed using linear mixed effects models using mixed function in the "afex" package (Singmann et al. 2020). The mixed function is a wrapper to the glmer function (for binary accuracy data) and the lmer function (for continuous response time data) in the "lme4" package (Bates et al. 2015) in R (R Core Team 2021). Fixed factors included speech type (clear and casual, sum coded [1, −1]), masker type (silent and masked, sum coded [1, −1]), and experimental half (first half and second half, sum coded [1, −1]). For both the accuracy and response time analyses, we used a backward-stepping approach with the random effects (Matuschek et al. 2017), starting with a maximal random effects structure. This structure, for both analyses, included random by-subject slopes and intercepts for masker type, speech type and their interaction as well as random by-sentence slopes and intercepts for speech type. Model comparisons, between the more complex and the simplified model, were conducted using the native *anova* function. We selected the more complex model if it was a significantly better fit to the data than the simplified model. For the response time models, we determined the best nonlinear transformation to apply to account for response time skew before stepping with random effects. Following recommendations from Lo and Andrews (2015), we found that an inverse Gaussian transformation was a better fit to the data than a gamma transformation. Model syntax and outputs for accuracy and response time are listed in Table 1.

Growth Curve Analysis • Pupil dilation was modeled using GCA using the "lme4" package (Bates et al. 2015) in R (R Core Team 2021). GCA is a type of multilevel regression that controls for collinearity by orthogonalizing the polynomial time terms thus enabling us to model pupil dilation changes over a defined temporal window (Mirman 2014). There is no universal standard for analyzing pupil data (for an elaborated discussion on various analysis approaches for pupil data, see Fink et al. 2023), though GCA is applied frequently for analyzing continuous pupillometry data in speech perception (McLaughlin & Van Engen 2020; McLaughlin et al. 2021b). For all GCAs, we first used a forward-stepping procedure to systematically test whether subsequent orthogonal polynomials and experimental variables of interest improved model fit. Model comparisons were conducted using the *anova* function. If a given polynomial did not improve model fit, we chose the simpler model with the best fitting lower order polynomial. All base models started with random by-subjects intercepts, which was then forward-stepped after the fixed polynomial terms were determined.

†Winn et al. (2018) found, in a systematic evaluation of different baseline window lengths (100 to 3000 msec), that the corrected pupil traces were essentially identical regardless of the baseline window length.

‡All baseline values were plotted for each participant for each trial to evaluate whether any systematic trends were present. There were no notable patterns in baseline values across the duration of the experiment or by any of the conditions of interest.

§One participant's responses indicated that they may have confused the labels but were otherwise able to determine whether the probe word was heard in the target sentence (accuracy of 36%). We determined that the binomial probability of scoring 0.36 and lower by chance was less than 5%. Thus, we inverted the participant's responses.

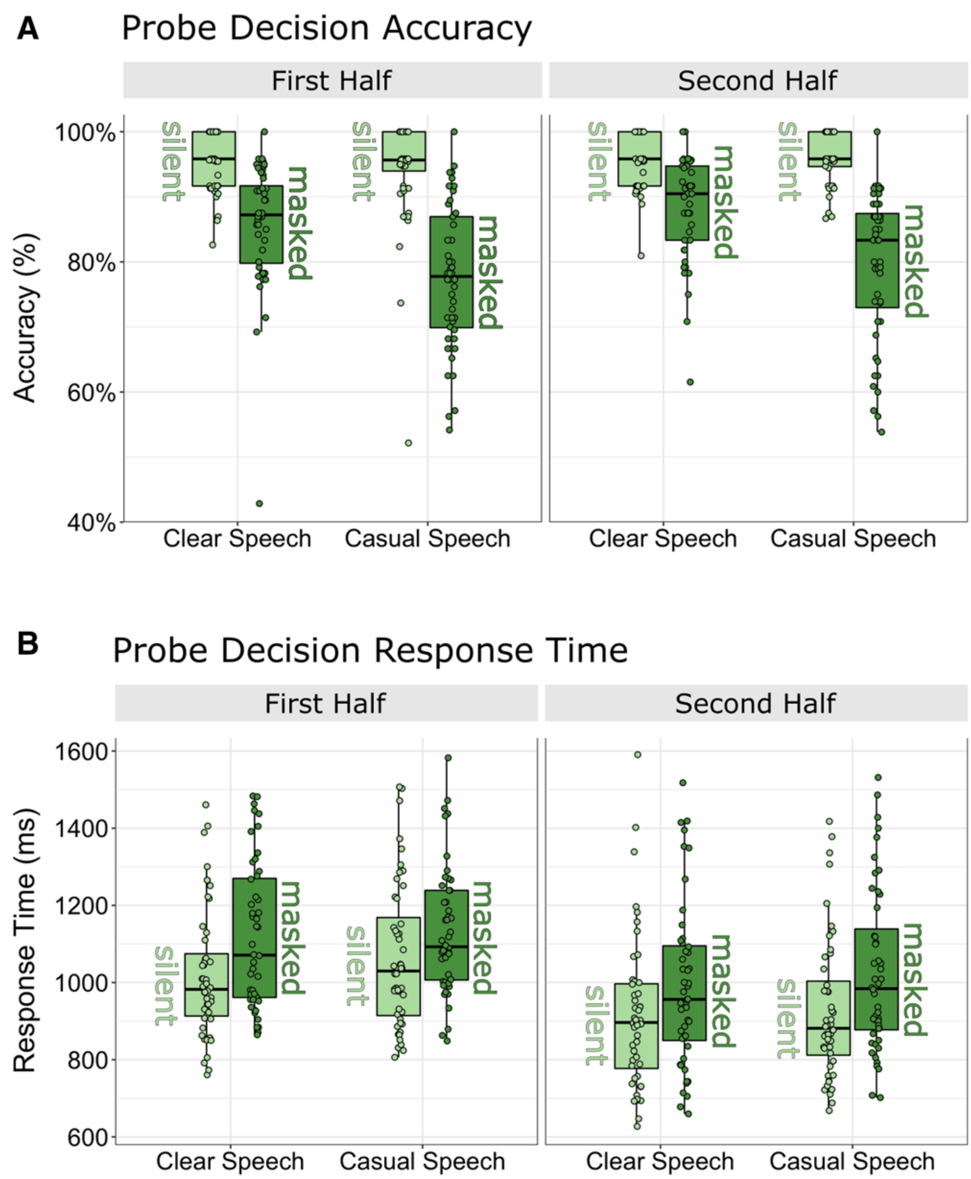

Fig. 2. Behavioral result plots for the probe word decision. A, Accuracy results, in percent correct, for the probe word decision as indicated by a box and whisker plot. Results are split on the *x* axis by speech type, with probe words following clearly spoken speech on the left and decisions following casually spoken speech on the right. Facets indicate which half of the experiment—first half in the left facet and second half is shown in the right facet. Trial type is marked by color and label: silent trials in light green and masked trials in dark green. Individual points indicate individual participant averages. B, Average response time, in msec, for the probe word decision. Plot follows convention of plot A, with speech type listed on the *x* axis and color indicating trial type (silent or masked). Individual points correspond to average individual participant response time.

RESULTS

Behavioral Results: Probe Decision

The full model fits and syntax are given in Table 1 for accuracy and response time of probe word decisions. For decision accuracy (see Fig. 2A), there were significant main effects of speech type ($p < 0.001$), masker type ($p < 0.001$), experiment half ($p < 0.001$), and a statistically significant interaction of speech type and masker type ($p < 0.010$). Accuracy was higher for probe word responses for unmasked target sentences than for masked target sentences. A main effect of speech type indicates that participants had higher accuracy for decisions following clearly spoken sentences than for casually spoken sentences, and the effect of experiment half suggests that participants were more accurate in the second half of the experiment than the first half, across all manipulations. To unpack the interaction of speech type and masker type, we used the package "emmeans" (Lenth et al. 2022). There was a significant effect of speech type in the masked condition ($p < 0.0001$), with responses less accurate for probes following casual speech ($M = 1.72$) than for clear speech ($M = 2.53$). Accuracy was similar, however, for probe decisions following unmasked speech ($p = 0.091$) for casually ($M = 3.93$) and clearly ($M = 4.23$) spoken target sentences. Results for response time (also shown in Table 1, bottom; plotted in Fig. 2B) show significant main effects of speech type (p

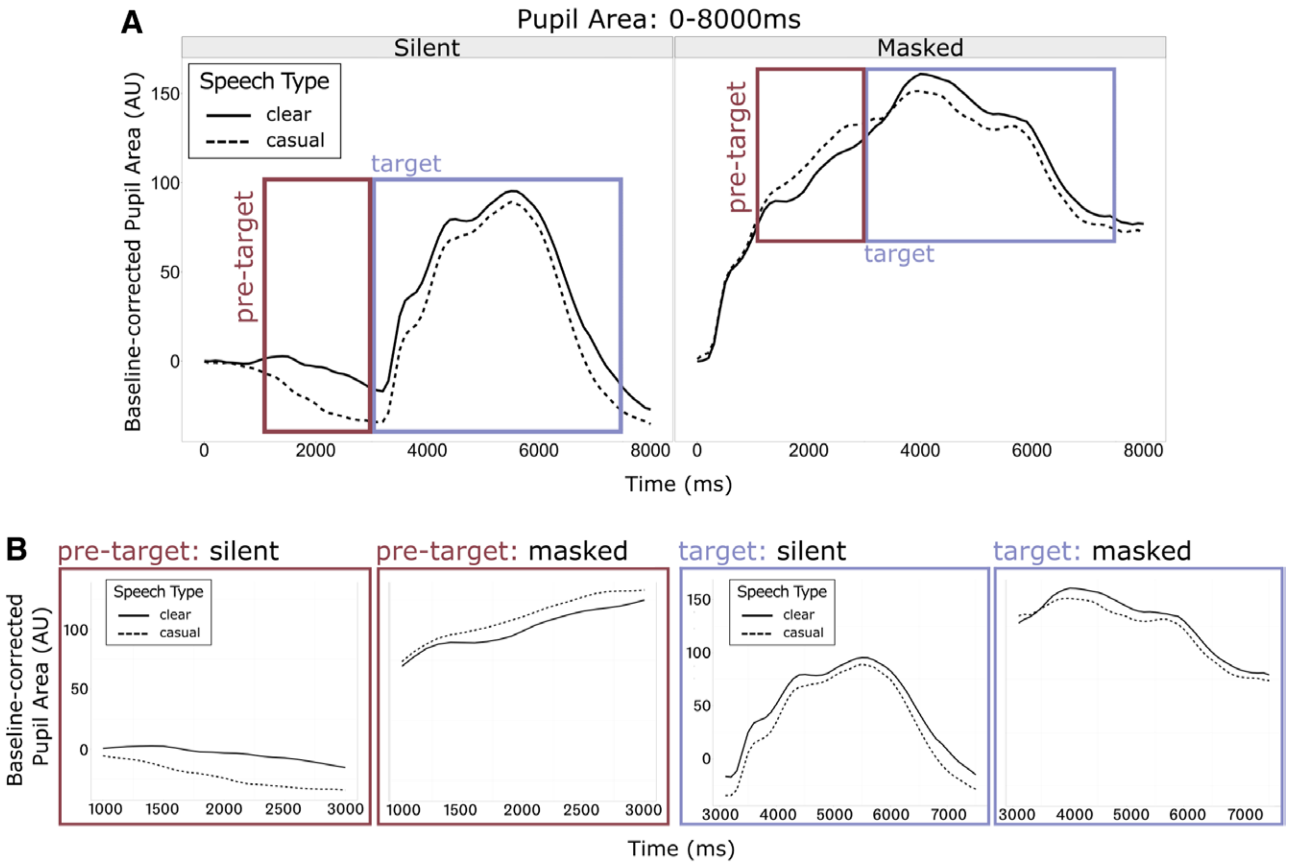

Fig. 3. Pupillary response plots. A, Plots showing the pupillary response over time. The baseline corrected pupillary response (pupil area) is plotted in arbitrary units along the *y* axis. Time, in msec, is indicated along the *x* axis, with the 0 timepoint indicating the start of the 3000 msec silent period or masked period before target sentence onset. The pupil trace concludes 8000 msec later. The lefthand plot is the aggregate for all silent trials while the righthand window is the aggregate for all masked trials. Solid lines correspond to pupil responses to clearly spoken sentences and dashed lines to casually spoken sentences. Drawn windows show where the pre-target windows (red boxes; analyzed post hoc) and the target sentence (purple boxes) windows fit in the larger pupillary response. B, Analysis windows broken down by trial type: silent or masked. The two leftmost panels are the pre-target windows, or pupillary responses before the onset of the target sentence. The two rightmost panels are the target sentence windows, or pupillary responses during the target sentence. As described earlier, solid lines are clear speech traces and dashed lines are casual speech traces.

TABLE 1. Probe word decision accuracy (top) and response time (bottom) results

Accuracy Model:Accuracy ~ Speech Type × Masker Type × Experiment Half + (1|Subject) + (1|Sentence Number)					
Fixed Effects	*β* Estimate	SE	*z* Value	*p*	Significance
Speech type	0.273	0.051	5.330	<0.001	***
Masker type	0.989	0.053	18.502	<0.001	***
Experiment half	−0.176	0.051	−3.427	<0.001	***
Speech type:masker type	−0.139	0.051	−2.721	0.0065	**
Speech type:experiment half	0.0492	0.051	0.960	0.337	
Condition:experiment half	−0.0226	0.051	−0.443	0.66	
Speech type:condition:experiment half	0.064	0.051	1.260	0.0201	
Response Time Model:Response_Time ~ Speech Type × Masker Type × Experiment Half + (Speech Type:Masker Type:Experiment Half|Subject) + (1|Sentence Number)					
Fixed Effects	*β* Estimate	SE	*t* Value	*p*	Significance
Speech type	−13.963	3.660	−3.815	<0.001	***
Masker type	−44.394	4.292	−10.344	<0.001	***
Experiment half	55.049	4.945	11.132	<0.001	***
Speech type:masker type	−0.504	4.603	−0.110	0.91	
Speech type:experiment half	−3.637	3.542	−1.027	0.305	
Condition:experiment half	0.969	3.667	0.265	0.791	
Speech type:condition:experiment half	−6.385	3.68	−1.651	0.0988	.

*Model syntax provided for both models. Significance indicated by asterisk number adhering to classic convention (*** for p < 0.001, ** for p < 0.01, * for p < 0.05, and . for p > 0.05).*

< 0.001), masker type ($p < 0.001$), and experiment half ($p < 0.001$), and no evidence for any two-way interactions (all, $p > 0.30$). These results suggest that participants responded faster to probes following unmasked target sentences, for clearly spoken target sentences, and in the second half of the experiment. A lack of evidence for an interaction suggests that there was not a compounding effect of masked, casually spoken target sentences, or experiment half on response time.

Pupillometry Results

Before we analyzed the pupil response during the target sentence window, we plotted the timecourse from trial onset (baselined before mask onset; see Fig. 3A). We noticed an interesting pattern immediately preceding the onset of the target sentence—pupil dilations differed by the upcoming target sentence speech type. As this window, which we will henceforth refer to as the "pre-target window," overlapped substantially with our a priori baseline window, we decided to shift our baseline to the beginning of the trial (as described in the Materials and Methods). Considering the potential theoretical implications of pre-target pupil dilations on adaptive speech perception (Aston-Jones & Cohen 2005), we ran an additional post hoc GCA on this pre-target window, which began 2000 msec before target sentence onset. Results below report separate analyses for pupil trajectories starting after the target sentence (target sentence GCA) and this 2000 msec window (pre-target GCA).

Target Sentence GCA • The target sentence window began 100 msec after target onset and continued for 4500 msec. The fixed effects of masker type, speech type, and experiment half were deviation coded to intuitively examine interactions and main effects: masker type (silent = [0.5], masked = [−0.5]), speech type (clear = [0.5], casual = [−0.5]), experiment half (first half = [0.5], second half = [−0.5]). We modeled up to the third polynomial (linear, quadratic, cubic; determined by visual inspection of the curves) based on improved model fits in a forward-stepping procedure. The random effects structure included random slopes and intercepts for participant-by-factor interactions as well as random slopes and intercepts for sentence number (i.e., stimuli). To characterize any emergent interactions, we then adopted a different coding scheme to examine the changes in curve structure. Keeping one factor as deviation coded, we then cycled through a treatment coding scheme for both levels of the factor of interest. Put plainly, if an interaction of masker type and speech type was indicated in the omnibus model, we treatment coded masker type such that each level was set as the reference level (i.e., silent = 0, masked = 1, and vice versa) and examined the effects of speech type. We then visualized the estimates to understand how the curves changed depending on condition level. Model output and final fits for the target sentence curves are shown in Table 2 and Figures 4A, B, respectively.

There were main effects of the linear ($\beta = -94.24$, SE = 20.423, $p < 0.001$), quadratic ($\beta = -167.75$, SE = 17.505, $p < 0.001$), and cubic ($\beta = 21.84$, SE = 7.888, $p < 0.01$) time terms, indicating that their inclusion improved model fit. For our effects of interest, there was a significant main effect of masker type at the intercept ($\beta = -81.27$, SE = 11.157, $p < 0.001$) reflecting greater overall pupil dilation, similar to area under the curve (Winn et al. 2015), for masked target sentences compared with unmasked. In addition, there were significant interactions of masker type on the linear ($\beta = 130.84$, SE = 7.117, $p < 0.001$) and quadratic ($\beta = -178.82$, SE = 7.116, $p < 0.001$) time terms, and a marginally significant interaction with the cubic term ($p = 0.075$). Dilations during masked target sentences peak earlier, higher, and have a flatter, broader peak compared with in silence. For experiment half, there was a significant main effect ($\beta = 29.541$, SE = 7.494, $p < 0.001$), such that there was a larger overall pupil dilation for the first half of the experiment compared with the second half.

Turning now to effects that involve "speech type," we found a small, marginally significant main effect of speech type such that there was slightly greater overall dilation for clear speech than for casual speech ($\beta = 9.64$, SE = 5.166, $p = 0.068$), with no further interactions with time terms. Further, there was a significant interaction of masker type and speech type at the

TABLE 2. Growth curve results for target sentence window

Target Sentence Model: Pupil ~ (Poly1 + Poly2 + Poly3) × Masker Type × Speech Type + Experiment Half + (1 + Poly1 + Poly2 + Poly3 + Masker Type + Speech Type + Experiment Half|Subject) + (1|Sentence Number)

Fixed Effects	β Estimate	SE	*t* Value	*p*	Significance
Intercept	83.25	13.852	6.01	<0.001	***
Poly1	−94.24	20.423	−4.62	<0.001	***
Poly2	−167.75	17.505	−9.58	<0.001	***
Poly3	21.84	7.888	2.77	<0.01	**
Masker type	−81.27	11.157	−7.28	<0.001	***
Speech type	9.64	5.166	1.87	0.0680	.
Experiment half	29.54	7.494	3.94	<0.001	***
Poly1: masker type	130.84	7.117	18.38	< 0.001	***
Poly2: masker type	−178.82	7.116	−25.13	<0.001	***
Poly3: masker type	−12.67	7.117	−1.78	0.0749	.
Poly1: speech type	2.93	7.12	0.411	0.681	
Poly2: speech type	2.52	7.119	0.354	0.723	
Poly3: speech type	5.23	7.118	0.734	0.463	
Masker type: speech type	6.66	2.127	3.133	<0.01	**
Poly1: masker type: speech type	−13.95	14.236	−0.98	0.327	
Poly2: masker type: speech type	44.49	14.234	3.126	<0.01	**
Poly3: masker type: speech type	−15.54	14.234	−1.092	0.275	

*Model syntax provided for both models. Significance indicated by asterisk number adhering to classic convention (*** for p < 0.001, ** for p < 0.01, * for p < 0.05, and . for p > 0.05). β estimates rounded to nearest tenth.*

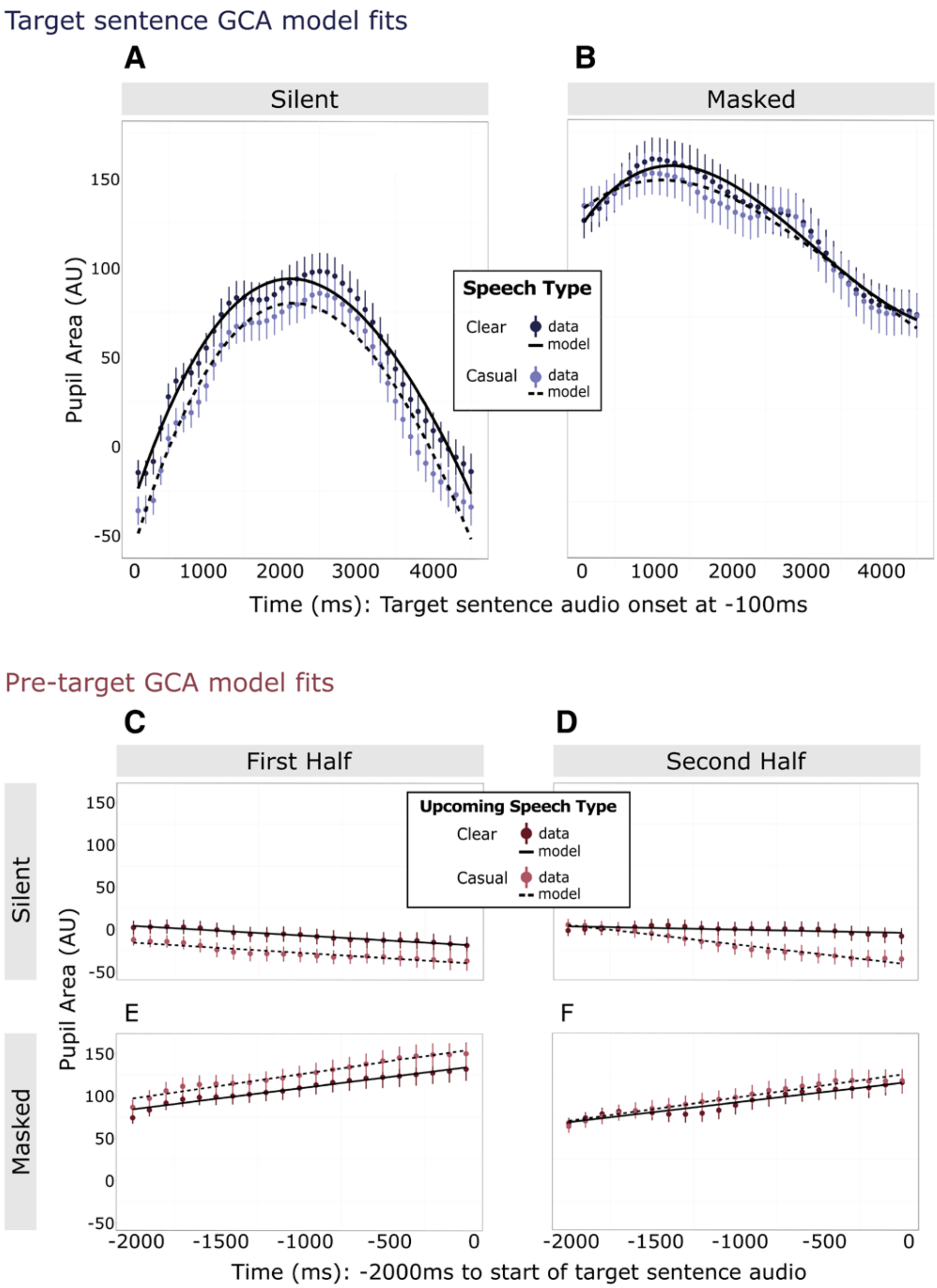

Fig. 4. All y axes show pupil area in arbitrary units. All x axes depict time, in msec, centered on the onset of the target sentence. A and B, Model fits for the target sentence window. The data curves are shown by dots with bisecting standard error lines with the dark purple dots corresponding to clearly spoken speech and light purple to casually spoken speech. Model fits are shown by lines: solid lines for clear speech and dashed lines for casual speech. A, Model fit for the target sentence during silent trials. B, Model fit for the target sentence during masked trials. C–F, Model fits for the pre-target window. Dark red dots correspond to clearly spoken speech and light red dots correspond to casually spoken speech. Model fit depiction follows the same convention as with the target sentence window: solid lines for clear speech and dashed lines for casual speech. C, Model fit for pre-target for silent trials in the first half of the experiment (blocks 1 to 8). D, Model fit for pre-target during silent trials in the second half of the experiment (blocks 9 to 16). E, Model fit for pre-target during masked trials in the first half of the experiment. F, Model fit for the pre-target during masked trials in the second half of the experiment. GCA indicates growth curve analysis.

intercept ($\beta = 6.66$, SE = 2,127, $p < 0.01$) and with the quadratic time term ($\beta = 44.49$, SE = 14.234, $p < 0.01$). We unpacked the interaction by running two submodels with a treatment coding scheme to examine the simple effects of speech type at each level of masker type. In silence, we saw a significant simple effect of speech type at the intercept ($\beta = 12.97$, $p < 0.05$) and an interaction with the quadratic term ($\beta = 24.770$, $p < 0.05$). Simply, clear speech has a larger and broader peak than casual speech when presented in silence. For masked speech, we found no evidence for a simple effect of speech type at the intercept ($p = 0.237$) and a marginally significant interaction ($p = 0.051$) with the quadratic term. Further examination revealed a similar pattern to that in silence; casual speech has a broader dilation curve than clear speech. Despite the marginal effect in noise,

TABLE 3. Growth curve results for pre-target window

Pre-Target Model: Pupil ~ (Poly1) × Masker Type × Speech Type × Experiment Half + (1 + Poly1 + Masker Type + Speech Type + Experiment Half|Subject)					
Fixed Effects	β Estimate	SE	t Value	p	Significance
Intercept	46.64	7.550	6.140	<0.001	***
Poly1	20.65	8.523	2.422	<0.05	*
Masker type	−119.21	9.729	−12.252	<0.001	***
Speech type	4.45	4.252	1.047	0.30	
Experiment half	7.551	5.554	1.360	0.18	
Poly1:masker type	−108.43	5.368	−20.198	<0.001	***
Poly1:speech type	4.18	5.370	0.778	0.437	
Masker type:speech type	28.462	2.349	12.117	<0.001	***
Poly1:experiment half	3.817	5.373	0.710	0.477	
Masker type:experiment half	−31.215	2.353	−13.264	<0.001	***
Speech type:experiment half	−1.091	2.357	−0.463	0.644	
Poly1:masker type:speech type	29.06	10.738	2.706	<0.01	**
Poly1:masker type:experiment half	6.710	10.744	0.625	0.532	
Poly1:speech type:experiment half	−21.93	10.746	−2.041	<0.05	*
Masker type:speech type:experiment half	19.42	4.706	4.126	<0.001	***
Poly1:masker type:speech type:experiment half	−54.79	21.488	−2.550	<0.05	*

*Model syntax provided for both models. Significance indicated by asterisk number adhering to classic convention (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$). β estimates rounded to nearest tenth.*

the interaction of masker and speech type was primarily driven by dilation differences that emerged in the silent condition. Notably, the effects of masking noise on pupil response were far larger than the more subtle effects of speech type, an issue we return to in the Discussion.

Pre-target GCA • The pre-target window began 2000 msec before target sentence onset and continued until target onset. As with the target sentence GCA, we included fixed effects of masker type, speech type, and experiment half. The fixed effects were coded identically to those for the target sentence GCA. Only the linear polynomial improved model fit, though we did examine whether the quadratic polynomial improved fit. It did not. For any interactions that emerged in the omnibus model, we followed the same treatment coding and subsequent visualization procedure as used with the target sentence GCA.

The final model output and syntax is shown in Table 3, and visualization of model fit superimposed on the pupil curves is shown in Figures 4C–F. As this was an unplanned post hoc analysis, we had no expectations for how pre-target anticipatory pupil dilation might change based on condition, upcoming speech type, or experiment half. For our fixed effects of interest, there were main effects of masker type at the intercept ($\beta = -119.21$, SE = 9.729, $p < 0.001$) and linear term ($\beta = -108.43$, SE = 5.368, $p < 0.001$). The model captured a larger overall dilation for masked trials than for silent ones. In addition, the model described the negatively sloped curve for silent trials and a steep, positive slope for masked trials. There was no evidence of a main effect of speech type at the intercept ($\beta = 4.45$, SE = 4.252, $p = 0.30$) and no detectable interaction with the linear term ($p = 0.44$). Last, there was no main effect of experiment half at either the intercept or linear term (all, $p > 0.18$). There were multiple 2-way interactions between our effects of interest (see Table 3), but we will focus on the significant 3-way interaction between masker type, speech type, and experiment half on the intercept ($\beta = 19.42$, SE = 4.706, $p < 0.001$) and linear terms ($\beta = -54.79$, SE = 21.488, $p < 0.05$).

We followed up on the interaction by running four submodels with a treatment coding scheme for masker type and experiment half to examine the simple effects of speech type at each level. For the first half of the silent trials there was a significant effect at the intercept ($p < 0.001$) but not at the linear term ($p = 0.58$) reflecting a larger overall pupil dilation for clear speech than for casual speech. For the second half silent trials, there were significant effects at the intercept ($p < 0.01$) and linear term ($p < 0.001$). Graphing beta estimates revealed an overall greater pupil dilation for clear speech than casual, as in the first half, but also a steeper negative slope for casual speech than for clear. For masked trials, the dilation pattern flipped. In the first half, there was a significant effect only at the intercept ($p < 0.01$) with greater overall dilation for casual speech than for clear. In the second half of the experiment this pattern collapsed, such that there were no significant effects of speech type at the intercept or at the linear term (both, $p > 0.20$).

DISCUSSION

To our knowledge, this study is among the first to use pupillometry to examine the perception of clear and casual speech; two speech types commonly encountered in everyday life. We investigated whether pupillometry could capture differences in listening effort while perceiving clear and casual speech in two background conditions: silent and masked. Concerning the effect of masking, we predicted overall greater dilations during masked speech due to reduced intelligibility of the target sentence; an extraordinarily robust finding (see Zekveld et al. 2010; for review, Zekveld et al., 2018). Initially adopting the view that casual speech is a type of degraded input, and thus less intelligible compared with clear speech (Mattys et al. 2012), paired with a highly consistent pattern in the literature showing greater pupil dilations for speech that is perceived as more demanding (for a recent review, see Van Engen & McLaughlin 2018), we expected greater dilations during casual speech compared with clear speech. This prediction was bolstered by findings that our participants were somewhat slower and less accurate in detecting probe words from casual compared with clear speech. Further, we anticipated that this effect—greater dilations for casual speech—would be consistent across both background conditions. During the target window following the onset of the sentence, we found that (1) pupil dilations were larger, earlier, and persisted longer for masked trials than for silent trials (matching our prediction), and (2) dilations were, surprisingly, slightly larger and lasted longer for clear speech compared with casual speech in silent trials but not in masked trials. Last, in a series of post hoc analyses, we discovered systematic differences in pre-target pupil area that we believe have interesting theoretical implications for the study of speech perception as well as important methodological considerations.

Effect of Noise Masking on the Pupillary Response

Larger pupil size is strongly associated with decreased target speech intelligibility, a finding replicated in the present study. Results indicated not only a larger peak pupil dilation for masked trials, but also a peak that lasts longer (i.e., broader) and peaks earlier than responses to silent trials. Although the theoretical implications of pupil dilation morphology for spoken sentence reception are not yet clear (Winn et al. 2015; Wendt et al. 2018), the effect of intelligibility and peak pupil dilation is remarkably consistent (for reviews, see Van Engen & McLaughlin 2018; Zekveld et al. 2018).

We found a significantly larger peak pupil area in noise compared with silence. These results are taken to reflect greater cognitive effort, or listening effort, to perceive speech in acoustically challenging environments (Zekveld et al. 2010; Winn et al. 2015; Kadem et al. 2020). It is interesting that, the relationship between speech intelligibility and pupil dilation is not linear but U-shaped. In prior work, when participants listened to spoken sentences with intelligibility ranging from 0 to 99%, the highest peak pupil dilations corresponded to conditions of intermediate intelligibility and dropped for very easy and very difficult conditions (Zekveld & Kramer 2014). Intelligibility, in large part, is experimentally modulated by introducing external noise at varying SNRs. This result replicates other findings also showing a larger pupillary response when listeners are confronted with background noise (for review, see Van Engen & McLaughlin 2018). The larger pupil size, together with the significantly reduced accuracy and increased response time suggests that listeners in the present study indeed found that the noise level manipulated increased stimulus difficulty. Although comforting to replicate a highly consistent finding, our primary goal was to investigate whether perception of clear and casual speech result in different pupillary response patterns and if those responses are consistent across both levels of background condition—questions which we explore in the sections later.

Effects of Experiment Half on the Pupillary Response

We also found evidence, both in the pupil and behavioral results, of sensitivity to experiment half. There are two likely, and unfortunately entangled, interpretations for the changes in behavior over time. The larger pupil dilations observed in the first half compared to the second half is consistent with accounts of listening-related fatigue and pupil response (McGarrigle et al. 2017; McLaughlin et al. 2022), or, alternatively, to the fact that sentences were repeated (albeit in a different noise condition) in the second half, potentially leading to less-effortful recognition. Familiarity with the stimuli could have boosted accuracy and lowered reaction time, and potentially resulted in attenuated pupil dilations. In addition, fatigue effects in pupillometry are well-documented (McGarrigle et al. 2017) which could also explain the smaller dilations in the second half of the experiment. Future studies that systematically manipulate these two factors—stimuli repetition and length of the experiment—will be necessary to tease apart how each factor independently affects pupil dilation during speech perception.

Greater Pupil Dilations for Clear Speech Than for Casual Speech

There is consistent behavioral evidence of a “clear speech benefit” when listening to speech in noise—there is improved recognition accuracy for clear speech compared with casual speech (Gagné et al. 1995; Ferguson & Kewley-Port 2002). Further, in a pupillometry study that compared naturally produced speech that varied in rate (faster speaking rate is a characteristic trait of casual speech), there were larger pupil dilations for faster speech than for slower speech in young adults (Koch & Janse 2016). Our prediction, then, was intuitive. We expected larger pupil dilation during perception of casual speech compared with clear speech, as casual speech likely requires greater listening effort to accurately perceive.

Our results show the inverse, with the effect of speech type on pupil dilation corresponding to larger dilations for clearly spoken target sentences, although this result only approached significance ($p = 0.068$), a result which should be interpreted with due caution. This can be seen in the upper panels of Figure 3, where pupil size is consistently greater during clear speech perception than during casual speech perception, though this effect is small in magnitude. In line with behavioral findings from the present study and others, we predicted that effects of clear speech would be amplified in noise. This prediction, too, was not borne out. Instead, while we found a significant interaction between speech type and masking condition, upon teasing apart the simple effects we found that the interaction was largely driven by a significant simple effect of speech type when presented in silence rather than noise These results were surprising, as previous work near-unanimously reports that less intelligible speech evokes larger pupil dilation (for recent reviews, see Van Engen & McLaughlin 2018; Zekveld et al. 2018). We offer several suggestions as to why our pattern differs from the established thinking that as speech becomes less intelligible (to a point), pupil size increases.

Task-evoked pupillary responses are thought to measure listening effort during speech perception tasks (Zekveld et al. 2018). Pichora-Fuller et al. (2016) in establishing the Framework for Understanding Effortful Listening, offers a domain-general definition of effort as the “deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task.” As already discussed, an obstacle that obscures the auditory signal is environmental noise, which reliably increases listening effort and therefore pupil size (Zekveld & Kramer 2014; Ohlenforst et al. 2018). Obstacles can also sprout from the auditory signal itself, such as distortion due to vocoding which is accompanied by increased pupil size relative to unmodified speech (Winn et al. 2015; Colby & McMurray 2021). Thus, the obstacles to speech perception in the present study are twofold. An external obstacle of environmental noise, induced by the multi-talker masker, and an obstacle in the acoustic signal itself—clear speech is thought to optimize intelligibility compared with casual speech (Smiljanić & Bradlow 2009). While our results appear to contradict these findings, keep in mind that the intelligibility of sentences in the present study was extremely high in silence (as rated by transcription during norming). It may be that the magnitude of the signal demand between speech types was nearly identical, which allowed other sources, such as listener motivation, to play a significant role in shaping the pattern of pupil responses.

Pichora-Fuller et al.'s (2016) framework qualifies their general definition of listening effort by adding that effort is modulated by "our motivation to achieve goals and attain rewards of personal and/or social value." In the laboratory, listener goals are highly constrained by the task. In the present study, the goal was to correctly make two-alternative forced choices at the end of each trial, which in turn motivated our listeners to closely attend the auditory sentences. The issue of motivation in listening effort is a sticky one, as it is routinely mentioned as having a significant effect on effort (Peelle 2018; Francis & Love 2020), but is largely omitted from discussions of pupillometric speech perception studies. We suggest that our pattern of results, with greater pupil dilations (i.e., greater listening effort) during clearly spoken sentences than for casual ones, reflects participant motivation to maximize goal achievement—that is, in the present study, correct word identification. This position is consistent with our behavioral findings, as participants had greater accuracy for responses following clear speech even though each stimulus category (clear versus casual) was rated as equally intelligible (in silence, by transcription) in a previous study (Xie & Myers 2018). Perhaps the clear speech benefit is not only a function of the reduced acoustic-phonetic demand but also reflects the prioritization of the speech signal that will result in the highest likelihood of accurate perception—in other words, the listener's motivation to understand.

In a guide to best practices for pupillometric speech perception research, Winn (2018) briefly introduced how the field of behavioral economics may provide insight into the question of listener motivation, or willingness, factors into pupil dilation. Behavioral economics is concerned with mapping effort-reward trade-offs in decision making and is directly relevant to our current findings. Paradigms originating from behavioral economics have been used in cognitive studies to investigate effort and motivation (Chiew & Braver 2013; Hopstaken et al. 2015). For example, increased rewards during the late stages of a task can result in dramatic re-engagement, as indexed by much larger pupil dilations after a steady decrease (Hopstaken et al. 2015). Further, a speech perception study found that older adults chose easier listening conditions (i.e., less background noise) even if they received less reward than if they chose harder, but more rewarding, listening conditions as compared with younger adults (McLaughlin et al. 2021a). This may suggest that listener motivation and effort expenditure are linked, though the exact nature of that relationship depends on a variety of factors including age, working memory, previous task performance, fatigue, available reward, and others. Moreover, Westbrook and Braver (2015), in applying neuroeconomics to cognitive effort, posit that effort is not simply a function of the demands of the task, in this case of the present study, intelligibility, but is emergent from a complex interaction of motivation, attention, arousal, and task. We endorse this perspective and suggest that the current results provide evidence that allocation of cognitive resources is highly sensitive to task conditions. Our task may have pushed listeners to concentrate more resources toward the signal most likely to result in accurate perception—clear speech.

We acknowledge that the effect sizes reported here are small. As such, these findings are subject to the same cautious interpretation of any small effects. Although it is difficult to attribute these small effects to any one source, it may be that the optimal SNR that allows the maximal effects of speech clarity to emerge is not the +5 dB noise level we selected. A thorough examination of various SNRs using these same stimuli and masker would be of great benefit to our understanding of how sources of noise (external and internal to the speech signal) intersect with effort, as other studies have done using different materials (Koelewijn et al. 2012; Ohlenforst et al. 2018). Future experimentation will be concerned with replicating these results with meaningful sentences and with populations beyond young university students. Despite these clear limitations, we believe these results are compelling if only as a deviation from the current pattern in the literature and hope that pupillometry continues to be used as a tool to examine how people process speech.

Pre-target Pupil Dilation: Theoretical Implications and Methodological Cautions

Differences in evoked pupil dilation between conditions and speech types were not limited to the processing of target sentence speech signal itself but emerged approximately 2000 msec before target sentence onset. We adopt the phrase pre-target pupil dilation to refer to the pupil dilations that begin before the target sentence audio in the present study, though note that the literature refers to this window as prestimulus (Hutchinson et al. 2020) or anticipation (Seropian et al. 2022). In a series of post hoc GCAs, we found that background condition (i.e., silence or masked), speech type (i.e., clear or casual), and experiment half modulated the magnitude and direction of pre-target pupil dilation. In silent blocks, pupil responses were greater before clear targets than for casual target sentences in both halves, though in the second half there was a steeper negatively sloped dilation preceding casual speech. The opposite pattern appeared in masked blocks, with a positively sloped "ramping up" effect before both speech types and for both halves, and pupil size preceding casually spoken targets were larger than responses before clearly spoken targets in the first half of the experiment. Although we did not design our experiment to examine pre-target pupil changes, we believe that the blocked nature of our task, paired with the consistent temporal elements of our trials, allowed systematic differences in pupil dilation to emerge before sentence onset, and suggest that participants were able to, implicitly or explicitly, use regularities in the trial structure to anticipate the listening demand of the upcoming sentence.

We are not the first to report on how certain experiment structures, in tasks examining listening effort, lead to the emergence of potential "anticipatory" effects (Seropian et al. 2022). Despite studies reporting pre-target dilations, there is little consensus of what pre-target dilation might reflect, especially in speech perception. We aim to connect literature outside of speech perception to move towards understanding, and in doing so hope we encourage future experimentation on the pre-target dilation effect. We then conclude with a cautionary note on subtractive baseline correction during continuous speech pupillometry tasks. Notably, when stimuli are fully randomized over the entire experiment, issues of baseline pupil size drift and differences in anticipatory responses are minimized. However, any design involving blocked stimuli (as in our study) or where effects are predicted to evolve over time (e.g., in studies investigating learning, fatigue, or adaptation) must grapple with these issues.

The implications of pre-target pupil dilation have been an intriguing area of study, especially in the domains of visual processing and cognitive control, for the past 20 years. One potential framework to interpret these results comes from Aston-Jones and Cohen (2005) Adaptive Gain Theory. This theory connects the functioning of the locus-coeruleus norepinephrine system (thought to control task-evoked pupillary responses; Joshi et al. 2016) to effortful processing. Simply, they suggest two processing modes that evoke different patterns of task-evoked (termed, "phasic") and tonic, or baseline, pupil dilation. In the transient active state, when the task requires high effort to acquire an achievable reward, the system adopts an exploitative strategy characterized by low prestimulation pupil dilation and high task-evoked pupil dilation. Conversely, when the task does not require effort and/or the reward (or goal) is very difficult to obtain, strategies shift to exploration during the prestimulation window. Thus, prestimulus pupil responses are large as different strategies are considered and selected while task-evoked dilations are low. This pattern of prestimulus to task-evoked pupil responses has been found in studies of visual processing (Gilzenrat et al. 2010; Irons et al. 2017; Hutchison et al. 2020), cognitive control (Chiew & Braver 2013), and auditory sequence predictability (Milne et al. 2021).

The broad strokes of Adaptive Gain Theory (Aston-Jones & Cohen 2005) appear similar to the pre-target dilations for masked trials. However, we interpret the current findings using Adaptive Gain Theory with extreme caution. It would be inaccurate to directly compare the patterns in tonic activity described by Adaptive Gain Theory with our pre-target window, as participants were already likely in a "phasic" firing state (Joshi & Gold 2020) in the masked condition because they were actively hearing the multi-talker babble before target sentence onset. Instead of drawing direct comparisons, we instead consider the possibility of participants oscillating between a state of strategy selection and strategy implementation during the pre-target and target sentence windows. When listeners anticipated an upcoming casually spoken sentence, pupil dilation was large, potentially cycling through strategies to maximize the chance of accurate perception. However, the failure to adopt the best listening strategy is associated with smaller task-evoked dilation; a pattern observed during the perception of casual speech compared with clear speech. We suggest that perhaps listeners adopted a perceptual strategy to optimize performance for clearly spoken sentences early on but struggled to select the best strategy to perceive casual speech. This is but one tentative interpretation of the observed pattern of results. A challenge remains to probe how exploration and exploitation modes emerge in humans, in speech perception specifically, and their relation to different neuronal firing patterns in the locus coeruleus.

As robust as prestimulus dilations are in relation to task difficulty and reward attainability in the visual domain, they reflect more than one construct. In the present study, we found an interaction of pre-target dilation with experiment half. Here, the effect of experiment half can be interpreted as either accounting for stimulus familiarity (because sentences were repeated across trial types in the two halves), or as a proxy for task-related fatigue. Behavioral data better supports a familiarity account: participants were faster and more accurate in the second half of the experiment. Nonetheless, the effects of fatigue on pupil responses during listening are well-documented in the literature and cannot be ruled out. In a recent study, Alhanbali et al. (2021) examined pre-target pupil dilation and self-reported measures of fatigue during a speech in noise task (single digits) with older adults. They found that, when difficulty was kept constant, larger pupil areas were associated with less fatigue and better performance. Pre-target dilation has also been interpreted as reflecting differing degrees of arousal during listening, as participants with moderate hearing loss have had larger pre-target dilations compared with normal hearing controls at the beginning of the experiment, which then decrease (perhaps due to fatigue) as the experiment continued (Ayasse & Wingfield 2020). A similar pattern was found when comparing people with aphasia and healthy controls (Chapman & Hallowell 2015). Pre-target pupil dilation is of theoretical interest to the study of listening effort during speech perception, as it may be a physiological marker of fatigue and difficulty (Zekveld et al. 2010), goal-based decision making (Aston-Jones & Cohen 2005), or all of the above. This is a promising area for future research, especially when considering the potential clinical implications of clear versus casual speech perception (Liu et al. 2004) and speech perception more broadly (Seropian et al. 2022).

As intriguing as the theoretical implications of pre-target pupil dilation is, its existence may threaten the integrity of task-evoked pupil responses. Subtractive baseline correction is the current standard for pupillometric data preprocessing due to the linear nature of the task-evoked pupil response and a desire to control for tonic pupil fluctuations (Reilly et al. 2019). Further, subtractive baselining has been found to be more resistant to bias compared with other baselining procedures (Mathôt et al. 2018). Typically, a subtractive baselining procedure requires establishing a baseline window immediately preceding event onset, taking the median pupil area within that window, and then subtracting all other values from the baseline value (Geller et al. 2020). We support using baseline correction as a preprocessing step, as was used in this present study. We add a note of caution with regard to where in the trial the baseline window is defined in terms of the whole trial timecourse, especially in blocked designs.

The present study used a blocked design, where we alternated speaking style and switched background conditions (noise or masked) every other block. Blocking is not uncommon in experiments investigating variable SNRs on listening effort (Koelewijn et al. 2011, 2014; Zekveld & Kramer 2014; Wendt et al. 2017, 2018; Ohlenforst et al. 2018). A blocked task structure (by all conditions of interest) paired with a highly consistent trial structure makes an environment ripe for prediction, and, we argue, the emergence of pre-target pupil divergences as listeners implicitly or explicitly learn the structure of the study; potentially launching anticipatory resources toward the upcoming input. This is not a problem for most experiments, particularly those that randomize at least one condition (Brown et al. 2020; McLaughlin & Van Engen 2020). Essentially, we (unintentionally) created the perfect storm for systematic pre-target dilation: (1) blocked all conditions of interest and (2) small task-evoked differences. The problem lies in how a sufficiently large pre-target dilation during the baseline window may result in an inversion of the task-evoked pupil response; a potentially catastrophic bias that emerged during the analysis of the current results. As a check against these kinds of errors, we therefore recommend plotting the entire trial timecourse using an early baseline window (see Figure in Supplemental Digital Content 1, http://links.lww.com/EANDH/B244 for graphic depiction of the effects of subtractive baseline correction, and see Peelle & Van Engen 2021 for a discussion of the contribution of researcher choices to outcomes in pupil response data).

CONCLUSION

Pupillometry can capture differences in listening effort between perception of clear and casual speech. During sentence perception, listeners appear to invest more effort while perceiving clear speech; an effect that was larger when there was no competing background noise. At first glance, this finding is at odds with research showing that as the speech signal becomes less intelligible, pupil dilation (and assumed listening effort) increases (Zekveld et al. 2018). We believe that listener motivation plays a crucial role in the allocation of effort during tasks (Peelle 2018), and as such participants in this study optimized their performance by dedicating more cognitive resources to the perception of clearly spoken sentences. An exciting avenue for future research is to untangle the complicated snarl of listening effort, motivation, and task to understand how pupil dilations relate to each construct. One notable ecological limitation of the present study is the nature of the stimuli themselves—sentences devoid of meaning. Using meaningless sentences certainly reduces our ecological explanatory power, but in doing so we isolated listening effort associated with processing phonetic and lexical ambiguity from the incoming signal. A logical next step includes using meaningful clear and casual productions to see if the effect reported here extends to more naturalistic stimuli. Differences in pupil dynamics also occurred before the onset of the target sentence itself. We offer a tentative interpretation of pre-target pupil dilation relating to effort/engagement trade-offs (Aston-Jones & Cohen 2005). Further experimentation, with carefully designed tasks meant to look specifically at pre-target dilation during speech perception, will be necessary to investigate the intriguing theoretical implications of this effect. Last, the presence of pre-target dilation complicates current standards in baselining procedures during pupillometry preprocessing (Reilly et al. 2019). By adding an early baseline and visualizing the entire timecourse we believe that biasing effects can be mitigated. Pupillometry has shown to be an extraordinarily sensitive method to detect differences in listening effort to highly intelligible speech at multiple timepoints.

ACKNOWLEDGMENTS

The authors thank Sahil Luthra, Pam Fuhrmeister, and the undergraduate members of the Language and Brain Lab for their extraordinary devotion to this project.

This research was supported by NIH R01 DC013064 (E.B.M., PI).

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIH NIDCD) awarded to the E.B.M.: NIH R01 DC013064. The funding organization had no role in conducting the experiment, analyzing the data, or preparing the article for publication.

H.R.M. designed the experiment, with advice from E.B.M. Data collection was done by H.R.M. and C.G. Data analysis and graphs were completed by H.R.M. The initial draft of the article was done by H.R.M., E.B.M, and C.G. provided comments and feedback for the final version. All authors reviewed and approved the final article for submission.

All data and analysis scripts are publicly available at: https://osf.io/4ft9c.

The authors have no conflicts of interest to disclose.

Address for correspondence: Hannah Mechtenberg, Department of Psychological Sciences, University of Connecticut, Storrs Mansfield, CT 06269, USA. E-mail: hannah.mechtenberg@uconn.edu

OPEN PRACTICES

This manuscript qualifies for an Open Data badge. The materials and data have been made publically available athttps://osf.io/4ft9c/?view_only=3223598c46f74174bcf763dc5c405406. More information about the Open Practices Badges can be found at https://journals.lww.com/ear-hearing/pages/default.aspx.

Received November 10, 2022; accepted September 1, 2023; published online ahead of print October 26, 2023.

REFERENCES

Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., Millman, R. E. (2021). Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy. *Int J Audiol*, *60*, 762–772.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu Rev Neurosci*, *28*, 403–450.

Ayasse, N. D., & Wingfield, A. (2020). Anticipatory baseline pupil diameter is sensitive to differences in hearing thresholds. *Front Psychol*, *10*, 2947.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *ArXiv:1406.5823 [Stat]* http://arxiv.org/abs/1406.5823

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull*, *91*, 276–292.

Boersma, P. & Weenink, D. (2019). Praat: Doing phonetics by computer [computer program].

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Commun*, *14*, 325–337.

Bradlow, A. R., Torretta, G. M., Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun*, *20*, 255–272.

Brown, V. A., McLaughlin, D. J., Strand, J. F., Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Q J Exp Psychol*, *73*, 1431–1443.

Chapman, L. R., & Hallowell, B. (2015). A novel pupillometric method for indexing word difficulty in individuals with and without aphasia. *J Speech Lang Hear Res*, *58*, 1508–1520.

Chiew, K., & Braver, T. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Front Psychol*, *4,* 15. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00015

Colby, S., & McMurray, B. (2021). Cognitive and physiological measures of listening effort during degraded speech perception: Relating dual-task and pupillometry paradigms. *J Speech Lang Hear Res*, *64*, 3627–3652.

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear*, *35*, 600–610.

Dryden, A., Allen, H. A., Henshaw, H., Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends Hear*, *21*, 2331216517744675.

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.

Fink, L., Simola, J., Tavano, A., Lange, E., Wallot, S., Laeng, B. (2023). From pre-processing to advanced dynamic modeling of pupil data. *Behav Res Methods*.https://doi.org/10.3758/s13428-023-02098-1

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognit Sci*, *11*, e1514.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K. G., Masterson, V. M. (1995). Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech. *Volta Rev.*, *97*, 33–51.

Geller, J., Winn, M. B., Mahr, T., Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behav Res Methods*, *52*, 2232–2255.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci*, *10*, 252–269.

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *J Acoust Soc Am*, *116*, 3108–3118.

Heald, S., & Nusbaum, H. (2014). Speech perception as an active cognitive process. *Front Syst Neurosci*, *8,* 35. https://www.frontiersin.org/article/10.3389/fnsys.2014.00035

Holube, I., Haeder, K., Imbery, C., Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends Hear*, *20*, 2331216516667734.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*, 305–315.

Hutchison, K. A., Moffitt, C. C., Hart, K., Hood, A. V. B., Watson, J. M., Marchak, F. M. (2020). Measuring task set preparation versus mind wandering using pupillometry. *J Exp Psychol Learn Mem Cogn*, *46*, 280–295.

Irons, J. L., Jeon, M., Leber, A. B. (2017). Pre-target pupil dilation and the preparatory control of attention. *PLoS One*, *12*, e0188787.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *J Cogn Neurosci*, *23*, 1587–1596.

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends Cogn Sci*, *24*, 466–480.

Joshi, S., Li, Y., Kalwani, R. M., Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*, 221–234.

Kadem, M., Herrmann, B., Rodd, J. M., Johnsrude, I. S. (2020). Pupil dilation is sensitive to semantic ambiguity and acoustic degradation. *Trends Hear*, *24*, 2331216520964068.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*, *116*, 2395–2405.

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *J Acoust Soc Am*, *139*, 1618–1636.

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hear Res*, *312*, 114–120.

Koelewijn, T., Zekveld, A. A., Festen, J. M., Kramer, S. (2011). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear*, *33*, 291–300.

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *Int J Otolaryngol*, *2012*, 1–11.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *J Acoust Soc Am*, *112*, 2165–2172.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am*, *115*, 362–378.

Książek, P., Zekveld, A. A., Wendt, D., Fiedler, L., Lunner, T., Kramer, S. E. (2021). Effect of speech-to-noise ratio and luminance on a range of current and potential pupil response measures to assess listening effort. *Trends Hear*, *25*, 233121652110093.

Kwon, K.-A., Shipley, R. J., Edirisinghe, M., Ezra, D. G., Rose, G., Best, S. M., Cameron, R. E. (2013). High-speed camera characterization of voluntary eye blinking kinematics. *J R Soc Interface*, *10*, 20130227.

Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *J Speech Lang Hear Res*, *56*, 1429–1440.

Lenth, R. V., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., Singmann, H. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means (1.8.2)*. https://CRAN.R-project.org/package=emmeans

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*, 2374–2383.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Front Psychol*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01171

Malsburg, T. v. d. (2015). *saccades: Detection of Fixations in Eye-Tracking Data (0.1-1).* https://CRAN.R-project.org/package=saccades

Mathôt, S., Fabius, J., Van Heusden, E., Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav Res Methods*, *50*, 94–106.

Mattys, S. L., Davis, M. H., Bradlow, A. R., Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang Cognit Process*, *27*, 953–978.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J Mem Lang*, *94*, 305–315.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*, 193–203.

McLaughlin, D. J., Braver, T. S., Peelle, J. E. (2021a). Measuring the subjective cost of listening effort using a discounting task. *J Speech Lang Hear Res*, *64*, 337–347.

McLaughlin, D. J., & Van Engen, K. J. (2020). Task-evoked pupil response for accurately recognized accented speech. *J Acoust Soc Am*, *147*, EL151–EL156.

McLaughlin, D. J., Zink, M. E., Gaunt, L., Reilly, J., Sommers, M. S., Van Engen, K. J., Peelle, J. E. (2022). Give me a break! Unavoidable fatigue effects in cognitive pupillometry. *Psychophysiology*, *60*, e14256.

McLaughlin, D. J., Zink, M. E., Gaunt, L., Brent Spehar, Van Engen, K. J., Sommers, M. S., Peelle, J. E. (2022). Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults. *Psychon Bull Rev, 29,* 268–280.

Milne, A., Zhao, S., Tampakaki, C., Bury, G., Chait, M. (2021). Sustained pupil responses are modulated by predictability of auditory sequences. *J Neurosci*, *41*, 6116–6127.

Mirman, D. (2014). Growth curve analysis: A hands-on tutorial on using multilevel regression to analyze time course data. In *Proceedings of the Annual Meeting of the Cognitive Science Society,* Vol. 36.

Moresi, S., Adam, J. J., Rijcken, J., Van Gerven, P. W. M., Kuipers, H., Jolles, J. (2008). Pupil dilation in response preparation. *Int J Psychophysiol*, *67*, 124–130.

Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Dev Psychol*, *41*, 352–362.

Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int J Audiol*, *52*, 433–441.

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear Res*, *365*, 90–99.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear Res*, *351*, 68–79.

Paulus, M., Hazan, V., Adank, P. (2020). The relationship between talker acoustics, intelligibility, and effort in degraded listening conditions. *J Acoust Soc Am*, *147*, 3348–3359.

Payton, K. L., Uchanski, R. M., Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am*, *95*, 1581–1592.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear*, *39*, 204–214.

Peelle, J. E., & Van Engen, K. J. (2021). Time stand still: Effects of temporal window selection on eye tracking analysis. *Collabra Psychol*, *7*, 25961.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *J Acoust Soc Am*, *23*, 148–148.

Picheny, M. A., Durlach, N. I., Braida, L. D. (1985). Speaking clearly for the hard of hearing I. *J Speech Lang Hear Res*, *28*, 96–103.

Picheny, M. A., Durlach, N. I., Braida, L. D. (1986). Speaking clearly for the hard of hearing II. *J Speech Lang Hear Res*, *29*, 434–446.

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *Int J Audiol*, *42*, 26–32.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear, 37,* 5S.

Porretta, V., & Tucker, B. V. (2019). Eyes wide open: Pupillary response to a foreign accent varying in intelligibility. *Front Commun*, *4,* 8.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Reilly, J., Kelly, A., Kim, S. H., Jett, S., Zuckerman, B. (2019). The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry. *Behav Res Methods*, *51*, 865–878.

Rodman, C., Moberly, A. C., Janse, E., Başkent, D., Tamati, T. N. (2020). The impact of speaking style on speech recognition in quiet and multi-talker babble in adult cochlear implant users. *J Acoust Soc Am*, *147*, 101–107.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7,* 31. https://www.frontiersin.org/article/10.3389/fnsys.2013.00031

Seropian, L., Ferschneider, M., Cholvy, F., Micheyl, C., Bidet-Caulet, A., Moulin, A. (2022). Comparing methods of analysis in pupillometry: Application to the assessment of listening effort in hearing-impaired patients. *Heliyon*, *8*, e09631.

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R., Christensen, R. H. B. (2020). *afex: Analysis of Factorial Experiments (0.26-0)*. https://CRAN.R-project.org/package=afex

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *J Acoust Soc Am*, *118*, 1677–1688.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Lang Linguist Compass*, *3*, 236–264.

Stanners, R. F., Coulter, M., Sweet, A. W., Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motiv Emot*, *3*, 319–340.

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychon Bull Rev*, *22*, 757–765.

van der Feest, S. V. H., Blanco, C. P., Smiljanic, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *J Phon*, *73*, 158–177.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev*, *25*, 2005–2015.

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hear Res*, *369*, 56–66.

Verney, S. P., Granholm, E., Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *Int J Psychophysiol*, *52*, 23–36.

Wendt, D., Hietkamp, R. K., Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear Hear*, *38*, 690–700.

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hear Res*, *369*, 67–78.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cogn Affect Behav Neurosci*, *15*, 395–415.

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. B., Wendt, D., Koelewijn, T., Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends Hear*, *22*, 2331216518800869.

Wu, Y-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., Oleson, J. (2018). Characteristics of Real-World Signal-to-noise Ratios and Speech Listening Situations of Older Adults with Mild-to-Moderate Hearing Loss. *Ear Hear, 39,* 293–304.

Xie, X., & Myers, E. (2018). Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: A comparison of clear and conversational speech. *J Cogn Neurosci*, *30*, 267–280.

Zekveld, A. A., Koelewijn, T., Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends Hear*, *22*, 2331216518777174.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry: Processing load across a wide range of listening conditions. *Psychophysiology*, *51*, 277–284.

Zekveld, A. A., Kramer, S. E., Festen, J. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear*, *31*, 480–490.