

DECEMBER 27 2023

Pupillometry reveals differences in cognitive demands of listening to face mask-attenuated speech

Sita Carraturo ; Drew J. McLaughlin ; Jonathan E. Peelle; Kristin J. Van Engen

 Check for updates

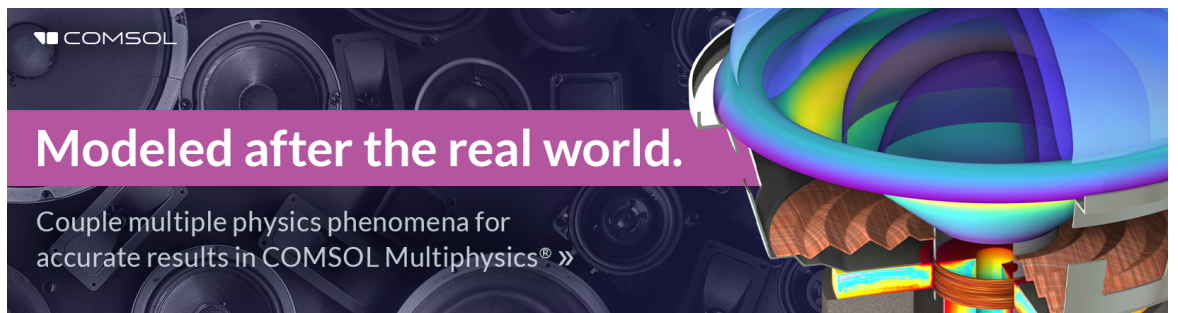
J. Acoust. Soc. Am. 154, 3973–3985 (2023)


<https://doi.org/10.1121/10.0023953>


View
Online


Export
Citation

 CrossMark



 **Modeled after the real world.**

Couple multiple physics phenomena for accurate results in COMSOL Multiphysics® »

Pupillometry reveals differences in cognitive demands of listening to face mask-attenuated speech

Sita Carraturo,^{1,a)}  Drew J. McLaughlin,²  Jonathan E. Peelle,^{3,b)} and Kristin J. Van Engen¹

¹Department of Psychological & Brain Sciences, Washington University in St. Louis, Saint Louis, Missouri 63130, USA

²Basque Center on Cognition, Brain and Language, San Sebastian, Basque Country 20009, Spain

³Department of Communication Sciences and Disorders, Northeastern University, Boston, Massachusetts 02115, USA

ABSTRACT:

Face masks offer essential protection but also interfere with speech communication. Here, audio-only sentences spoken through four types of masks were presented in noise to young adult listeners. Pupil dilation (an index of cognitive demand), intelligibility, and subjective effort and performance ratings were collected. Dilation increased in response to each mask relative to the no-mask condition and differed significantly where acoustic attenuation was most prominent. These results suggest that the acoustic impact of the mask drives not only the intelligibility of speech, but also the cognitive demands of listening. Subjective effort ratings reflected the same trends as the pupil data. © 2023 Acoustical Society of America. <https://doi.org/10.1121/10.0023953>

(Received 20 March 2023; revised 27 November 2023; accepted 29 November 2023; published online 27 December 2023)

[Editor: Benjamin V Tucker]

Pages: 3973–3985

I. INTRODUCTION

Although face masks can offer protection from airborne particles and viruses, they can also create significant challenges for spoken communication. Face masks introduce two forms of interference to communication: acoustic degradation (muffled speech) and visual occlusion (opaque masks prevent the listener from accessing a number of useful visual cues). However, because these aspects are affected concurrently, it is difficult to assess the degree to which acoustic challenge and visual occlusion independently contribute to making speech perception difficult. The focus of the current study is to isolate the acoustic effects of different types of face masks to measure how acoustic challenge affects the cognitive demands of comprehending face mask-attenuated speech.

Studies that have evaluated how face masks affect acoustic properties of speech tend to come to similar conclusions about the effects of different types of masks [e.g., Brown *et al.* (2021), Corey *et al.* (2020), and Magee *et al.* (2020)]. For example, Corey *et al.* (2020) measured the attenuation of speech through 12 types of masks. Across all the masks, sounds below 1 kHz were largely unaffected, but the higher frequency sounds (which are critical for speech perception) were attenuated to different degrees based on the material and construction of the mask. More specifically, results from Corey *et al.* show that surgical masks attenuate speech sounds the least, by about 4 dB. The attenuation by non-surgical cloth masks varied greatly (3–10 dB) depending on the specific fabric. Last, while they provide access to

visual cues, transparent masks may substantially attenuate speech frequencies (8 dB). Brown *et al.* (2021) measured the long-term average spectra of speech spoken through various masks and showed that acoustic speech was least affected by the surgical masks and most affected by cloth and transparent masks.

In the context of spoken communication, it is also important to consider not only the acoustic effects on speech, but how those acoustic changes affect listeners. Many studies have used speech intelligibility scores and shown that, for listeners with good hearing, face masks themselves do not greatly affect speech comprehension in quiet; rather, it seems that the combination of face masks and background noise is what reduces comprehension accuracy (Brown *et al.*, 2021; Magee *et al.*, 2020; Yi *et al.*, 2021). For example, across the four masks tested by Magee *et al.* (2020), intelligibility scores in quiet were not statistically different. Brown *et al.* (2021) compared intelligibility in quiet, moderate, and high levels of noise, and also found that although accuracy was at ceiling in quiet, it varied by mask in noisy conditions. Specifically, the intelligibility data in noisy conditions align with the acoustic data, with surgical masks yielding higher intelligibility scores than other masks.

While diminished intelligibility is a well-known outcome of listening in suboptimal conditions, there has been increased awareness that such conditions also impact the cognitive demands required to perform a listening task. In order to extract the intended message from a degraded speech signal, listeners typically need to engage additional cognitive resources [e.g., Rönnberg *et al.* (2013), Pichora-Fuller *et al.* (2016), and Peelle (2018)]. In much of the literature that has assessed the effort involved in understanding face mask-attenuated speech, researchers ask participants to

^{a)}Email: sita@wustl.edu

^{b)}Also at: Department of Psychology, Northeastern University, Boston, MA 02115, USA.

use subjective rating scales. Unsurprisingly, these data show that listeners perceive the task to be subjectively more effortful when the talker is wearing a mask (Brown *et al.*, 2021; Giovanelli *et al.*, 2021; Lee *et al.*, 2022). While these subjective reports tell us something about the listener's experience, the exact construct underlying these measures remains unclear, and subjective ratings do not always correlate with psychophysiological measures [e.g., Zekveld *et al.* (2010) and Strand *et al.* (2018)].

Pupillometry is an increasingly popular method of measuring cognitive demands during speech processing (for a review, see Van Engen and McLaughlin, 2018). Using pupillometry to measure the cognitive demands of a task is appealing because it is a physiological response that is not directly controlled by the subject, whereas a subjective rating measure can be sensitive to issues such as demand characteristics (Nichols and Maner, 2008). Neurologically, the task-evoked pupil response is a reflection of sympathetic and parasympathetic nervous system activity and projections from the locus coeruleus that relate to task engagement (Mathôt, 2018). Behaviorally, this response has been shown to be sensitive to task load, where pupil dilation is generally greater for more difficult tasks (e.g., Kahneman and Beatty, 1966). Many studies have specifically shown greater pupil dilation in response to increased difficulty for auditory tasks. For example, Winn *et al.* (2015) showed that pupil dilation increased with greater spectral degradation of the auditory signal and McLaughlin *et al.* (2022) demonstrated greater pupil dilation in response to key words with dense relative to more sparse neighborhood densities [for a comprehensive review, see Zekveld *et al.* (2018)]. Importantly, the pupil response is more sensitive to differences in task demands than speech intelligibility scores alone. Zekveld *et al.* (2010) had shown that pupil dilation increases as intelligibility decreases, but more recent studies have demonstrated a dissociation between intelligibility and cognitive demands—that is, differences in pupil dilation can be measured even when accuracy is unaffected (Koelewijn *et al.*, 2012; McLaughlin and Van Engen, 2020; Winn and Teece, 2021; McLaughlin *et al.*, 2022).

In the current study, we investigate how auditory speech produced through four types of face masks differentially affects the cognitive demands of spoken sentence processing in noise, using converging evidence from pupillometry, intelligibility, and subjective ratings. We isolate the auditory speech signal as opposed to providing an audiovisual signal, which allows us to address how the acoustic impact of different masks affect the cognitive demands of speech perception. Our design also addresses real-world instances in which face mask-attenuated speech is auditory-only (e.g., over the phone, communication with people who are blind, in workplace situations where eye-contact is not practical, etc.).

II. METHOD

The preregistered method, hypotheses, and analyses for this study can be found online (Carraturo *et al.*, 2023). The

experimental protocol was approved by the Institutional Review Board at Washington University in St. Louis.

A. Participants

The preregistered sample size for this study ($N = 54$) was based on related prior pupillometry studies that reported sample sizes of 50 or more [e.g., McLaughlin and Van Engen (2020)].

Participants were recruited from the Washington University in St. Louis Psychology Participant Pool and were compensated with course credit. All subjects were 18–22 years old ($M = 19.44$, $SD = 1.18$) and were native monolingual speakers of American English. Participants were screened ahead of time and deemed ineligible if they reported any known hearing or neurological impairments.

B. Materials

1. Speech stimuli

The audio-only speech recordings used in this study are the same as those used in an audiovisual context by Brown *et al.* (2021). The stimuli consisted of 156 sentences (150 experimental trials, 6 practice trials; all listed in the Appendix), each of which had four content words that would be scored (e.g., “The gray mouse ate the cheese”). The sentences were recorded by a female monolingual speaker of American English. The speaker recorded the sentences once while not wearing any mask, and four additional times while wearing each of the following types of masks: surgical, a fabric mask (60% cotton, 40% polyester; Safe Mate brand), the same fabric mask with a paper filter insert (brand unknown), and a “transparent” mask (a fabric mask with a clear plastic window; brand unknown).

As in Brown *et al.* (2021), each stimulus was mixed with pink noise, which had been generated in version 3.0.0 of Audacity® (Audacity Team, 2021). The noise was applied to the stimuli at 0 dB SNR, beginning 3500 ms prior to stimulus onset and ending 3000–5000 ms after stimulus offset. We chose a 0 dB SNR to mitigate both ceiling and floor effects of noise. We chose to focus this study on the effect of the face masks given *some* noise as opposed to the interaction of Mask Type and different noise levels. The no-mask stimuli were prepared twice: once with noise and once without noise, such that six conditions resulted: each of the four masks in noise, no-mask in noise, and no-mask in quiet. The long-term average spectra of these stimuli in quiet are plotted in Fig. 1, alongside images of the talker wearing each of the masks.

2. Subjective ratings

After each block, participants were asked to make two subjective ratings: the first was an effort rating, and the second was a performance rating. The wording for these ratings was derived from the NASA Task Load Index [NASA-TLX (Hart and Staveland, 1988)]. Specifically, participants were asked, “How hard did you have to work to accomplish your

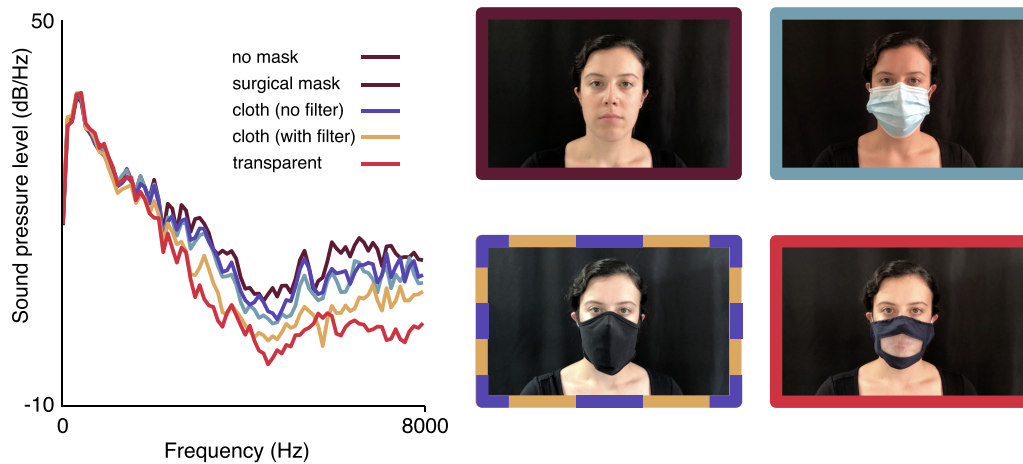


FIG. 1. (Color online) *Left*: long-term average spectra of stimuli in quiet. *Right*: speaker wearing each of the masks used. The cloth mask looked identical with and without a filter.

level of performance?” and “How successful were you in accomplishing what you were asked to do?” Participants were asked to give their ratings verbally on a scale of 1–100. For the effort rating, 1 indicated Low Effort and 100 indicated High Effort. For the performance rating, the scale was reverse-coded (as in the original NASA-TLX) such that 1 indicated Perfect and 100 indicated Failure.

C. Design

Each of the 150 sentences was randomly assigned to one of six lists (25 sentences per list), which were counter-balanced across the six conditions. Conditions were blocked, sentences within each block were randomized, and each block occurred equally in each presentation position (1st, 2nd, 3rd, 4th, 5th, or 6th). Each of the six counterbalances was presented to nine participants such that, across participants, each sentence was heard in each condition nine times (but a given participant only heard each sentence once).

There was one error in stimulus recording that affected one of the six counterbalances. Sentences 53 and 54 were “The three sisters watched the movie” and “The tiny kitten chased the mouse,” respectively. However, in the surgical mask recordings, sentence 53 was recorded as “The tiny kitten chased the mouse” and sentence 54 was recorded as “The small kitten chased the mouse.” As a result, the sentences that correspond with items 53 and 54 are distinct for nine participants. We report this error here for transparency but did not remove any of these trials from analyses.

D. Equipment

Pupil diameter was measured using an EyeLink 1000 Plus eye-tracker (SR Research, Ottawa, Canada) with a sampling rate of 500 Hz. The camera was set up according to EyeLink specifications in front of a computer monitor. Stimuli were presented on that monitor (screen resolution: 1024 × 768) using E-Prime 2.0 (Psychology Software Tools, Inc., 2012). Auditory stimuli were delivered binaurally through Sennheiser over-the-ear wired headphones.

E. Procedure

The experiment took place in a dimly lit and sound-attenuated room. Prior to beginning the experiment, participants read an information sheet and provided verbal consent. Participants were then seated at a desk facing a computer monitor. Participants heard verbal instructions for the task before assuming position in a desk-mounted chinrest. Once the participant was situated, the experimenter left the room and began calibrating the eye-tracker to track the participant’s right eye. After calibration, the main experimental task began. To minimize effects of metacognition, participants were not told that the speaker they would be listening to might be wearing a face mask. The instructions were simply that they would hear a female voice and that the task was to repeat what they heard to the best of their ability.

The experimental task is illustrated in Fig. 2. Participants read instructions on the screen and then pressed the spacebar to proceed. Participants were instructed to attend to the sentences and repeat them back while maintaining their gaze on a fixation cross on the monitor. The fixation cross (presented against a light gray background) was red during stimulus presentation, but otherwise blue. The cross-turned red 3500 ms before sentence onset. The red cross-was the participant’s cue to look at the cross, to reduce blinking as much as possible, and to listen to the sentence. The cross-remained red for 3500 ms after sentence offset, at which point it turned blue. This was the participant’s cue to repeat the sentence (making guesses for any words they were unsure of) and that they could blink freely. Participants’ verbal responses were recorded on an adjacent computer in the same room using Audacity® (Audacity Team, 2021). When the participant was ready to proceed onto the next sentence, they pressed the spacebar. At this time, the cross-remained blue for an additional 3000 ms delay, allowing for the pupil to return to baseline.

The first six trials were practice trials, and all participants heard the same six practice sentences (one for each condition) in random order. After the practice trials, the participants were able to ask the researcher questions and then proceed onto the first block by pressing the spacebar.

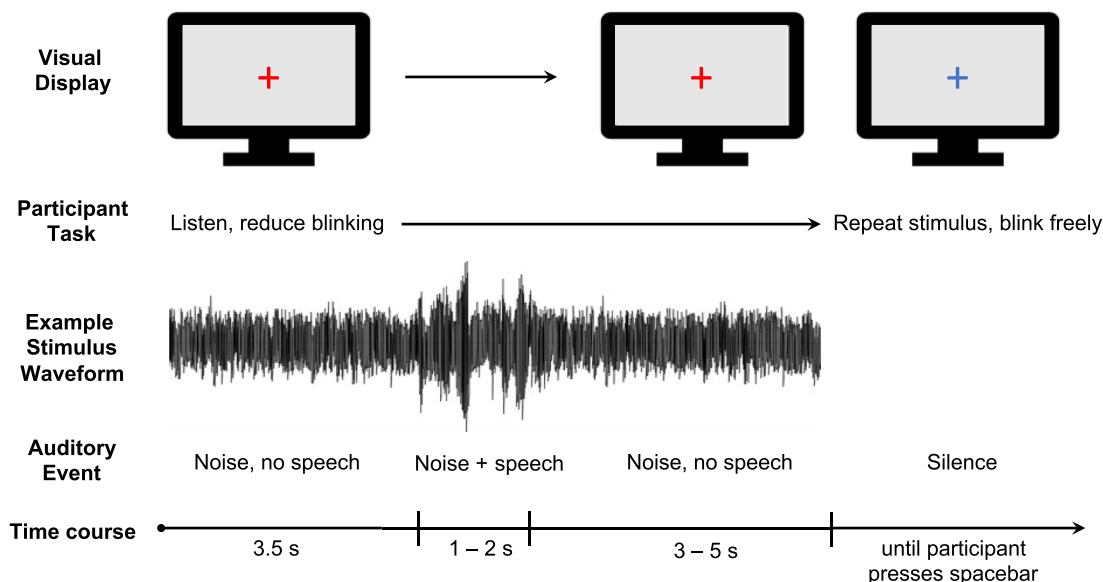


FIG. 2. (Color online) Task schematic. Data collection took place while the fixation cross was red.

After each block, participants were presented with the two subjective rating questions, one at a time. These questions appeared on the screen, but participants gave their responses aloud (also audio-recorded). Following these ratings, participants were given the chance to take a break (but did not get up or leave the room) before moving onto the next block. This task took approximately 50 min.

After the main experimental task, audiometric tests were conducted to collect pure-tone average hearing thresholds for each participant, across 1, 2, and 4 kHz in both ears (better ear was used in analyses). Participants then answered a brief demographic questionnaire that asked about age, language background, and education.

Finally, participants completed a working memory task. This was a shortened version of the reading span task from Oswald *et al.* (2015), also administered through E-Prime. In this task, participants read sentences on the screen and pressed a key to indicate whether or not they made sense. After each sentence, a letter of the alphabet appeared. After a set of sentence judgements and letters, participants were asked to recall the letters. There was a total of six trials with two each of list lengths 4, 5, and 6. This task took approximately 10 min.

The entire procedure took approximately 70 min.

III. DATA PREPROCESSING

Data were processed and analyzed in RStudio version 1.2.5042 for macOS (RStudio Team, 2020).

A. Pupillometric data

A series of custom R scripts employing functions from the gazeR (Geller *et al.*, 2020) package were used to prepare the pupil data for analyses. First, any trials in which 50% of the data were missing were excluded (this resulted in 2.1%

of total trials being excluded). Data were then processed to identify blinks, extend them 200 ms preceding and 100 ms following, and extrapolate across them linearly. The data were then smoothed via application of a five-point rolling average. Next, the pupil data were subtractive-baseline corrected (Reilly *et al.*, 2019). To achieve this, the pupil data were aligned at sentence onset. Then, data from the 500 ms preceding each sentence was averaged to calculate a per-trial baseline pupil dilation value. The resulting baseline value for each trial was then subtracted from the remaining pupil measures in that trial. The data were then down-sampled (to reduce computational load) from 500 to 50 Hz by time-binning the data.

The final step was to select a window of time for the growth curve analysis. We collapsed the data across all conditions and plotted it to select a window of analysis that maximized the amount of pupil data included but fit the constraints of a cubic shape. A starting point was selected at 200 ms (where 0 ms was stimulus onset), and the end point was selected at 2500 ms. Then we created a base growth curve model with only the linear, quadratic, and cubic polynomial terms as fixed effects and random intercepts by subject. This step validated that a growth curve model was appropriate for the pupil response curve. Based on visual inspection, we concluded that the model-predicted fit line adequately estimated the raw data.

B. Intelligibility

A single coder transcribed the participants' verbal responses. These transcriptions were then prepared for scoring via Autoscore (Borrie *et al.*, 2019) according to the website's instructions. Autoscore allows the user to select rules for the types of spelling deviations that would be accepted as correct answers. Of the available rules (which are

explained on the website), we selected the following: *root word*, *double letter*, *acceptable spell*, *tense*, and *plural rule*. Each target sentence had four key words that were scored such that accuracy scores per trial could be 0.0, 0.25, 0.5, 0.75, or 1.

IV. RESULTS

A. Deviations from the preregistration

The analyses below deviate from the preregistration in two ways: first, we intended to analyze all six conditions (no mask in quiet, no mask in noise, surgical mask in noise, cloth mask without filter in noise, cloth mask with filter in noise, and transparent mask in noise). However, upon plotting the raw pupillometry data by condition, it became apparent that the onset of background noise affected the baselining of the pupil diameter such that the “no mask in quiet” condition was not comparable to the five conditions with noise. Specifically, we interpret the baseline difference between the no mask in quiet condition and the other conditions as attributable to the onset of the background noise, which can elicit an early pupil response and dampen the stimulus-evoked pupil response. As such, all the analyses below exclude the “no mask in quiet” condition. The plotted raw data with the six conditions are available in the supplementary materials.¹

Second, [McLaughlin et al. \(2023\)](#) demonstrated that trial number (a proxy for time within the experiment for a given subject) has a large effect on model fit, as it captures the reduction in pupil response due to fatigue. We therefore include Trial as a fixed effect in all of the analyses below.

B. Pupil dilation and Mask Type

We used growth curve analysis ([Mirman, 2014](#)) to analyze the pupillometry data over time. We started by adding three orthogonal polynomials (linear, quadratic, and cubic) to the data frame via the `poly()` function in R, which centers each of the time predictors. Mixed effects models were run using the `lme4` package ([Bates et al., 2015](#)). We used log-likelihood model comparisons to determine the contribution of each fixed effect and interaction (Table I), and we used estimates obtained via model summaries to determine the direction of the effects.

We began with a more complex random effects structure that included random slopes for Mask Type, the polynomial terms, Trial, and their interactions. When that model did not converge, we reduced the complexity of the random effects structure, stepwise, by removing slopes for the interactions first and then the polynomial terms. The resulting random effect structure for all models included random intercepts by item and subject, and random slopes for Mask Type by item and subject. Fixed effects included the following: Linear Polynomial, Quadratic Polynomial, Cubic Polynomial, Mask Type (dummy-coded reference level: No Mask), and Trial. In the interaction model, fixed effects also included Mask Type × Linear Polynomial, Mask Type × Quadratic Polynomial, and Mask Type × Cubic Polynomial.

TABLE I. Model comparison results for pupil dilation and Mask Type analysis.

Effect	χ^2	df	p
Linear polynomial	4023.9	1	< 0.001
Quadratic polynomial	1037.8	1	< 0.001
Cubic polynomial	463.98	1	< 0.001
Mask Type	13.84	4	< 0.01
Trial	3115.1	1	< 0.001
Mask Type × Linear polynomial	3099.7	4	< 0.001
Mask Type × Quadratic polynomial	29.86	4	< 0.001
Mask Type × Cubic polynomial	10.264	4	< 0.05

All three polynomials improved model fit: Linear ($\chi^2 = 4023.9$, $DF = 1$, $p < 0.001$), Quadratic, ($\chi^2 = 1037.8$, $DF = 1$, $p < 0.001$), and Cubic ($\chi^2 = 463.98$, $DF = 1$, $p < 0.001$). The effect of Mask Type ($\chi^2 = 13.84$, $DF = 4$, $p < 0.01$) also improved model fit, as did Trial ($\chi^2 = 3115.1$, $DF = 1$, $p < 0.001$). Model estimates show that, relative to the No Mask condition, pupil dilation increased with each of the masks (ordered by size of estimate): Surgical Mask ($\beta = 10.37$, $SE = 23.98$, $t = 0.43$), Cloth Mask without Filter ($\beta = 20.19$, $SE = 2.07$, $t = 0.98$), Cloth Mask with Filter ($\beta = 63.42$, $SE = 22.77$, $t = 2.79$), and Transparent Mask ($\beta = 74.92$, $SE = 24.61$, $t = 3.04$).

Iteratively rotating the reference level in the base model and using the package `lmerTest` ([Kuznetsova et al., 2017](#)) to calculate p -values revealed that each of the increases in pupil dilation from the No Mask condition to the Surgical Mask condition, and from the Surgical Mask condition to the Cloth Mask without Filter condition were not statistically significant from one another ($p = 0.67$ and $p = 0.64$, respectively). The increase in pupil dilation from the Cloth Mask without Filter to the Cloth Mask with Filter was significant ($p < 0.05$). Last, the increase in pupil dilation from the Cloth Mask with Filter to the Transparent Mask was not statistically significant ($p = 0.57$). These differences are plotted in figure Fig. 3(a). A figure of model fits plotted over raw data are available in the supplementary materials.¹

Each of the model interactions also significantly improved model fit: Mask Type × Linear Polynomial ($\chi^2 = 3099.7$, $DF = 4$, $p < 0.001$); Mask Type × Quadratic Polynomial ($\chi^2 = 29.86$, $DF = 4$, $p < 0.001$); and Mask Type × Cubic Polynomial ($\chi^2 = 10.26$, $DF = 4$, $p < 0.05$). These significant interactions indicate that Mask Type had a significant effect on the rate of pupil dilation (linear), on the sharpness of the points of inflection in the curves (quadratic and cubic), and/or on the latency of the peak (cubic). Model estimates are available in supplementary materials.¹

C. Pupil dilation and intelligibility

A second set of analyses was conducted in which Intelligibility was added as a fixed effect. The most complex random effects structure that converged for these models included by-subject and by-item random intercepts, random slopes for Mask Type by item and subject, and random

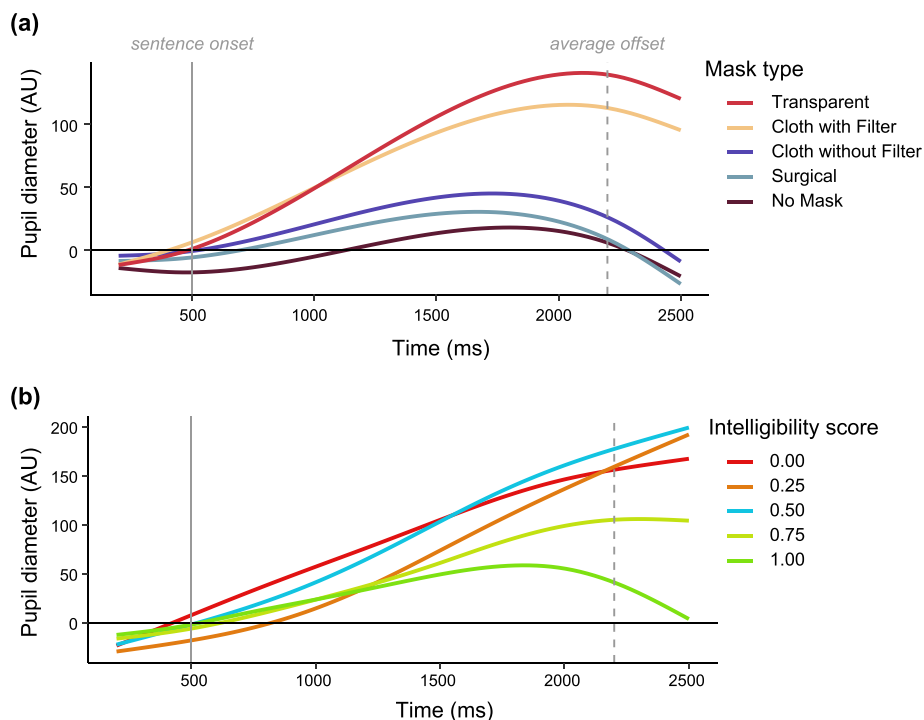


FIG. 3. (Color online) (a) Model fits of baseline-corrected pupil dilation are plotted as a function of time for each listening condition. Solid vertical line indicates stimulus onset; dashed vertical line indicates average stimulus offset. Pupil diameter is reported in arbitrary units (AU). (b) Model fits of pupil dilation are binned by the intelligibility accuracy for a given trial and plotted as a function of time.² Solid vertical line indicates stimulus onset; dashed vertical line indicates average stimulus offset. Pupil diameter is reported in arbitrary units (AU).

slopes for Intelligibility by subject. This second model allowed us to investigate the extent to which the pupil dilation is affected by the face masks, above and beyond what is captured by intelligibility. Log-likelihood model comparisons for this model are summarized in Table II.

Chi-squared and p-values for all three polynomials were equivalent to those reported in the previous model. The effect of Mask Type improved model fit ($\chi^2 = 14.82$, $DF = 4$, $p < 0.01$), as did Trial ($\chi^2 = 2957.7$, $DF = 1$, $p < 0.001$). Intelligibility did not improve model fit ($\chi^2 = 0.525$, $DF = 1$, $p = 0.47$).

Next, we tested whether the model interactions improved fit. The interaction between the effects of Intelligibility and Mask Type was significant ($\chi^2 = 435.14$, $DF = 4$, $p < 0.001$). Model estimates indicated that in the No Mask condition there was a trend toward larger pupil responses for more

intelligible materials ($\beta = 129.00$, $SE = 25.50$, $t = 5.06$), whereas for the Surgical Mask condition ($\beta = -178.00$, $SE = 13.80$, $t = -12.92$), Cloth with Filter condition ($\beta = -198.00$, $SE = 11.60$, $t = -17.03$), and Transparent condition ($\beta = -157.00$, $SE = 11.00$, $t = -14.22$) this trend was reversed; for the Cloth without Filter condition ($\beta = -52.80$, $SE = 13.00$, $t = -4.07$) this trend was reduced but not fully reversed. In other words, the general trend was toward larger pupil response for less intelligible materials [see Fig. 3(b)], as has been found in prior work (Zekveld *et al.*, 2011), with the exceptional case being the No Mask condition.

The interactions between the effects of Mask Type and the polynomials were, again, all significant: Mask Type \times Linear polynomial ($\chi^2 = 2176.3$, $DF = 4$, $p < 0.001$); Mask Type \times Quadratic polynomial ($\chi^2 = 45.62$, $DF = 4$, $p < 0.001$); and Mask Type \times Cubic polynomial ($\chi^2 = 14.46$, $DF = 4$, $p < 0.01$). Here, too, model estimates showed that, as compared to the No Mask condition, the rate of pupil dilation was steeper (linear) and the points of inflection were more pronounced (quadratic and cubic) in the Cloth with Filter and Transparent Mask conditions. For the Surgical and Cloth without Filter conditions, the rate of dilation was lesser and equal to the No Mask condition, respectively.

The interactions between Intelligibility and the polynomial terms also improved model fit: Intelligibility \times Linear polynomial ($\chi^2 = 669.38$, $DF = 1$, $p < 0.001$); Intelligibility \times Quadratic polynomial ($\chi^2 = 190.48$, $DF = 1$, $p < 0.001$); Intelligibility \times Cubic polynomial ($\chi^2 = 7.84$, $DF = 1$, $p < 0.01$). The nature of these interactions is such that, as intelligibility increased, the rate of pupil dilation decreased (Linear \times Intelligibility: $\beta = -539.1$, $SE = 20.83$, $t = -25.88$), and the pupil response became more curvilinear (Quadratic \times Intelligibility: $\beta = -287.5$, $SE = 20.83$, $t = -13.80$; Cubic

TABLE II. Model comparison results for pupil dilation and intelligibility analysis.

Effect	χ^2	df	p
Linear polynomial	4045.8	1	< 0.001
Quadratic polynomial	1043.4	1	< 0.001
Cubic polynomial	466.52	1	< 0.001
Mask Type	14.82	4	< 0.01
Trial	2957.7	1	< 0.001
Intelligibility	0.5253	1	0.47
Intelligibility \times Mask Type	435.13	4	< 0.001
Mask Type \times Linear polynomial	2176.3	4	< 0.001
Mask Type \times Quadratic polynomial	45.62	4	< 0.001
Mask Type \times Cubic polynomial	14.46	4	< 0.01
Intelligibility \times Linear polynomial	669.38	1	< 0.001
Intelligibility \times Quadratic polynomial	190.48	1	< 0.001
Intelligibility \times Cubic polynomial	7.84	1	< 0.01

Polynomial \times Intelligibility: $\beta = -583.3$, $SE = 20.83$, $t = -2.8$). These effects can be seen in Fig. 3(b).

The above analysis tested for the effects of both Mask Type and Intelligibility on pupil dilation. To further assess the independent contribution of Mask Type, we tested the effect of Mask Type for only the fully intelligible trials. Fixed effects included the following: Linear Polynomial, Quadratic Polynomial, Cubic Polynomial, Mask Type (dummy-coded reference level: No Mask), and Trial. The random effects structure included random intercepts by item and subject, and random slopes for Mask Type by item and subject. In the interaction model, fixed effects also included Mask Type \times Linear Polynomial, Mask Type \times Quadratic Polynomial, and Mask Type \times Cubic Polynomial.

All three polynomials improved model fit: Linear ($\chi^2 = 4023.9$, $DF = 1$, $p < 0.001$), Quadratic, ($\chi^2 = 1037.8$, $DF = 1$, $p < 0.001$), and Cubic ($\chi^2 = 463.98$, $DF = 1$, $p < 0.001$). The effect of Mask Type ($\chi^2 = 13.84$, $DF = 4$, $p < 0.01$) also improved model fit, as did Trial ($\chi^2 = 3115.1$, $DF = 1$, $p < 0.001$). Model estimates show that, relative to the No Mask condition, pupil dilation increased with each of the masks (ordered by size of estimate): Surgical Mask ($\beta = 10.37$, $SE = 23.98$, $t = 0.67$), Cloth Mask without Filter ($\beta = 20.19$, $SE = 20.71$, $t = 0.33$), Cloth Mask with Filter ($\beta = 63.42$, $SE = 22.77$, $t = 2.79$), and Transparent Mask ($\beta = 74.92$, $SE = 24.61$, $t = 3.04$).

As above, we iteratively rotated the reference level in the base model to make pairwise comparisons among the levels of Mask Type. This revealed that only the increase in dilation from the Cloth Mask without Filter condition to the Cloth Mask with Filter condition was significant ($p < 0.05$).

Each of the model interactions also significantly improved model fit: Mask Type \times Linear Polynomial ($\chi^2 = 3099.7$, $DF = 4$, $p < 0.001$); Mask Type \times Quadratic Polynomial ($\chi^2 = 29.86$, $DF = 4$, $p < 0.001$); and Mask Type \times Cubic Polynomial ($\chi^2 = 10.26$, $DF = 4$, $p < 0.05$).

D. Effect of Mask Type on Intelligibility

Next, we analyzed the effect of Mask Type on Intelligibility. For this analysis, each keyword in each trial was scored individually as correct (1) or incorrect (0), and as such, we used generalized linear mixed effects models.

For these models, participants and items (sentences) were included as random intercepts. Given that the listening condition was within-subjects and within-items, we also modeled by-participant random slopes and by-item random slopes. Comparison of models with and without Mask Type included as a fixed effects showed that Mask Type had a significant effect on Intelligibility ($\chi^2 = 109.48$, $DF = 4$, $p < 0.001$). Model estimates from the full model showed that all masks except the Surgical Mask ($\beta = -0.25$, $SE = 0.22$, $p = 0.32$) significantly reduced intelligibility relative to the No Mask condition (Cloth without Filter: $\beta = -0.65$, $SE = 0.24$, $p < 0.001$; Cloth with Filter: $\beta = -1.80$, $SE = 0.23$, $p < 0.001$; and Transparent: $\beta = -2.40$, $SE = 0.22$, $p < 0.001$).

Rotating the reference level of the model further showed that the Cloth Mask without Filter did not significantly reduce intelligibility relative to the Surgical Mask ($\beta = -0.40$, $SE = 0.24$, $p = 0.09$), that the Cloth Mask with Filter did significantly reduce intelligibility relative to the Cloth Mask without Filter ($\beta = -1.15$, $SE = 0.21$, $p < 0.001$), and that the Transparent Mask significantly reduced intelligibility relative to the Cloth Mask with Filter ($\beta = -0.60$, $SE = 0.15$, $p < 0.001$). These data are plotted in Fig. 4(a).

E. Subjective ratings

Next, we analyzed the data derived from the modified NASA-TLX questions using linear mixed effects models. These models included Mask Type as a fixed effect, and subject as a random effect. Because only one of each subjective rating was made per block, there were no by-item random effects. Model estimates are used to compare levels within Mask Type, and the package *lmerTest* was used to calculate the p -values. The analysis of the Subjective Effort

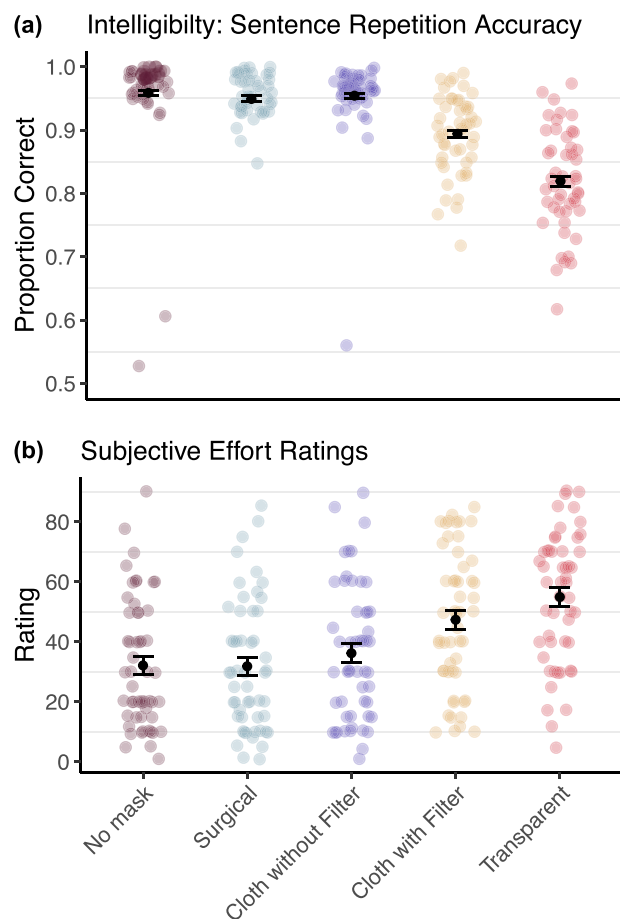


FIG. 4. (Color online) (a) Proportion of accurately repeated words plotted as a function of listening condition. Black dots represent means and bars denote standard error. Each colored point represents one participant's performance in that condition. (b) Subjective effort ratings per participant are plotted as a function of listening condition. Black dots indicate the mean and bars denote standard error. Each colored point represents one participant's rating in that condition.

ratings is below, but the analysis of Subjective Performance ratings can be found in supplementary materials.¹

1. Subjective effort ratings

The model comparisons showed that the effect of Mask Type improved model fit ($\chi^2 = 105.27$, $DF = 4$, $p < 0.001$). Effort ratings (on a scale of 1–100, where 1 was Low Effort and 100 was High Effort) by Mask Type are plotted in Fig. 4(b).

Relative to the No Mask condition, subjective effort ratings were only significantly different for the following three conditions: Cloth Mask without Filter ($\beta = 5.18$, $SE = 2.52$, $t = 2.06$, $p < 0.05$), Cloth Mask with Filter ($\beta = 16.18$, $SE = 2.51$, $t = 6.45$, $p < 0.001$), and Transparent Mask ($\beta = 23.74$, $SE = 2.51$, $t = 9.471$, $p < 0.001$). The reference level was then rotated to make comparisons among the mask types, and determine whether there were also

differences among these three conditions. This yielded the following results: Effort ratings were significantly higher for the Cloth Mask with Filter condition ($p < 0.001$) than for Cloth Mask without Filter condition, and were also significantly higher in the Transparent Mask condition ($p < 0.01$) than in the Cloth Mask with Filter condition.

We tested the correlations between subjective effort ratings and peak pupil diameters in each condition, but this exploratory analysis did not yield any significant results (all p 's > 0.05 ; see Fig. 5).

F. Individual differences measures

Last, we performed exploratory correlation analyses of peak pupil diameter in each mask condition with participants' (1) better ear pure tone averages (PTA) and (2) working memory scores from the reading span measure (RSPAN). Better ear PTAs correlated significantly with

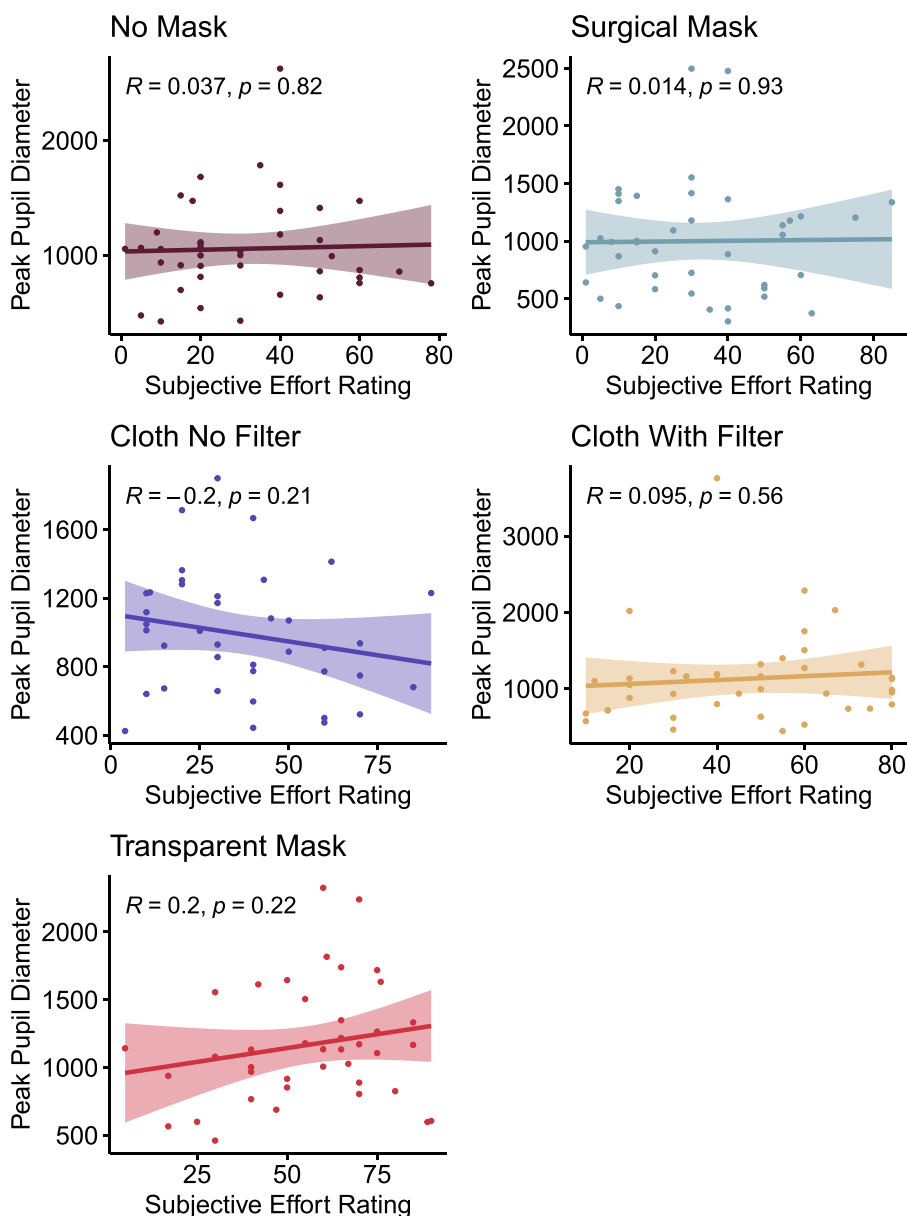


FIG. 5. (Color online) Correlations between Subjective Effort Ratings and Peak Pupil Diameter in each mask condition.

peak pupil diameter in only two conditions: No Mask ($R = 0.35, p < 0.05$), and Cloth Mask with Filter ($R = 0.57, p < 0.001$). In all other conditions, $p > 0.05$. These correlations are plotted in Fig. 6.

Reading span scores, on the other hand, did not significantly correlate with peak pupil diameter in any of the conditions (all $p > 0.05$). These correlations are plotted in Fig. 7.

V. DISCUSSION

In this study, we used pupillometry to assess how the resulting acoustics of different types of face masks differentially affect the cognitive demands of understanding speech in noise. The results indicate that the cognitive demands of listening increase significantly where acoustic information is most attenuated.

In our study, we used audio-only versions of the stimuli from Brown *et al.* (2021), and our results closely match

ones obtained with their audiovisual presentation. Brown *et al.* collected subjective reports of listening effort on an online study with older and younger adults. Their data, like ours, suggest that surgical masks result in the least exerted effort and transparent masks result in the greatest effort.

We also collected subjective effort ratings. Although those data reflected the same trends found in the pupillometry data (which masks were on average easier and which were harder), subjective ratings and peak pupil diameter were not significantly correlated. This issue is important within the greater discussion of the construct validity of *listening effort* (Strand *et al.*, 2018). Our data cannot clarify whether subjective ratings and pupil dilation are indexing the same underlying construct, but we do believe they contribute some useful insight into the unanswered question. In our design, participants provided one subjective rating per condition; on the other hand, the eye-tracker measured the

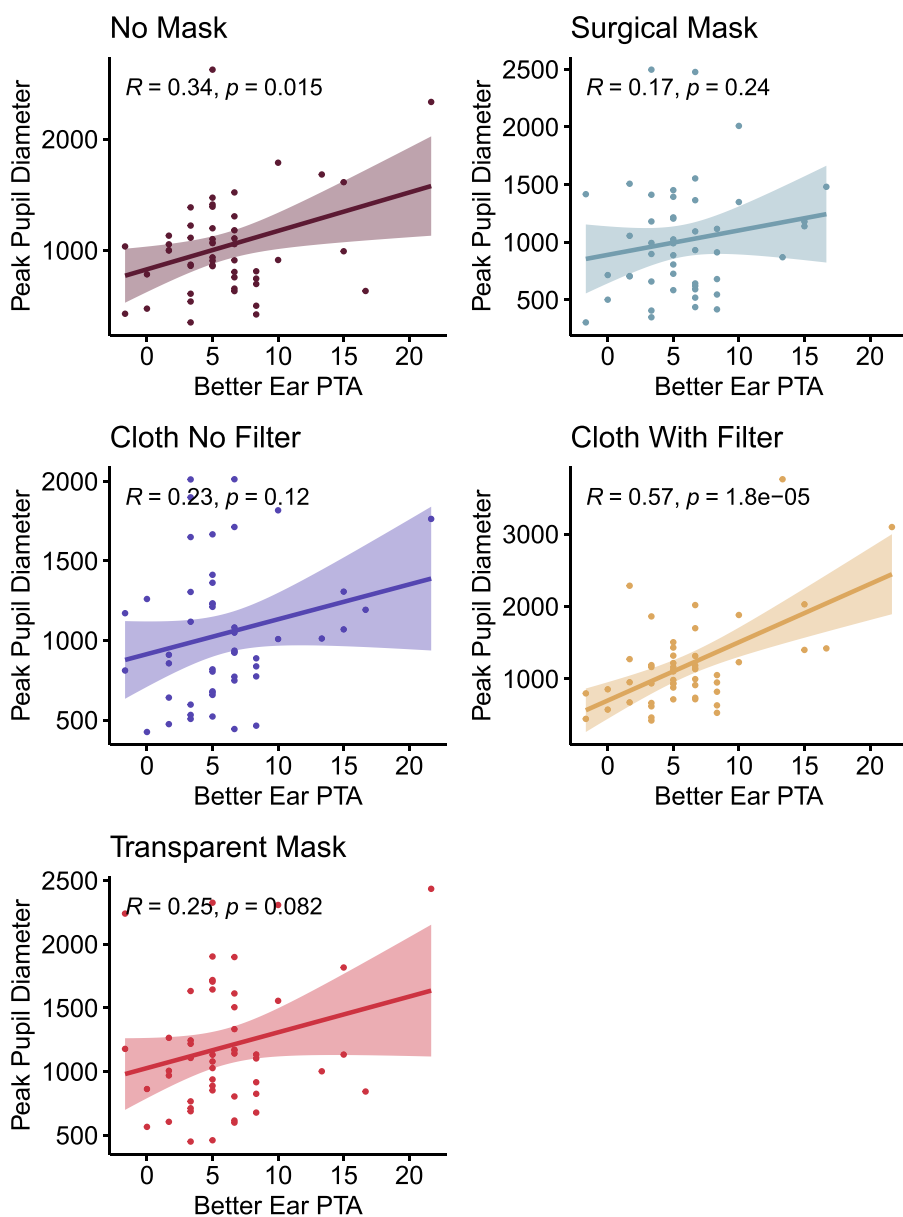


FIG. 6. (Color online) Correlations between Better Ear Pure Tone Average (PTA) and Peak Pupil Diameter in each mask condition.

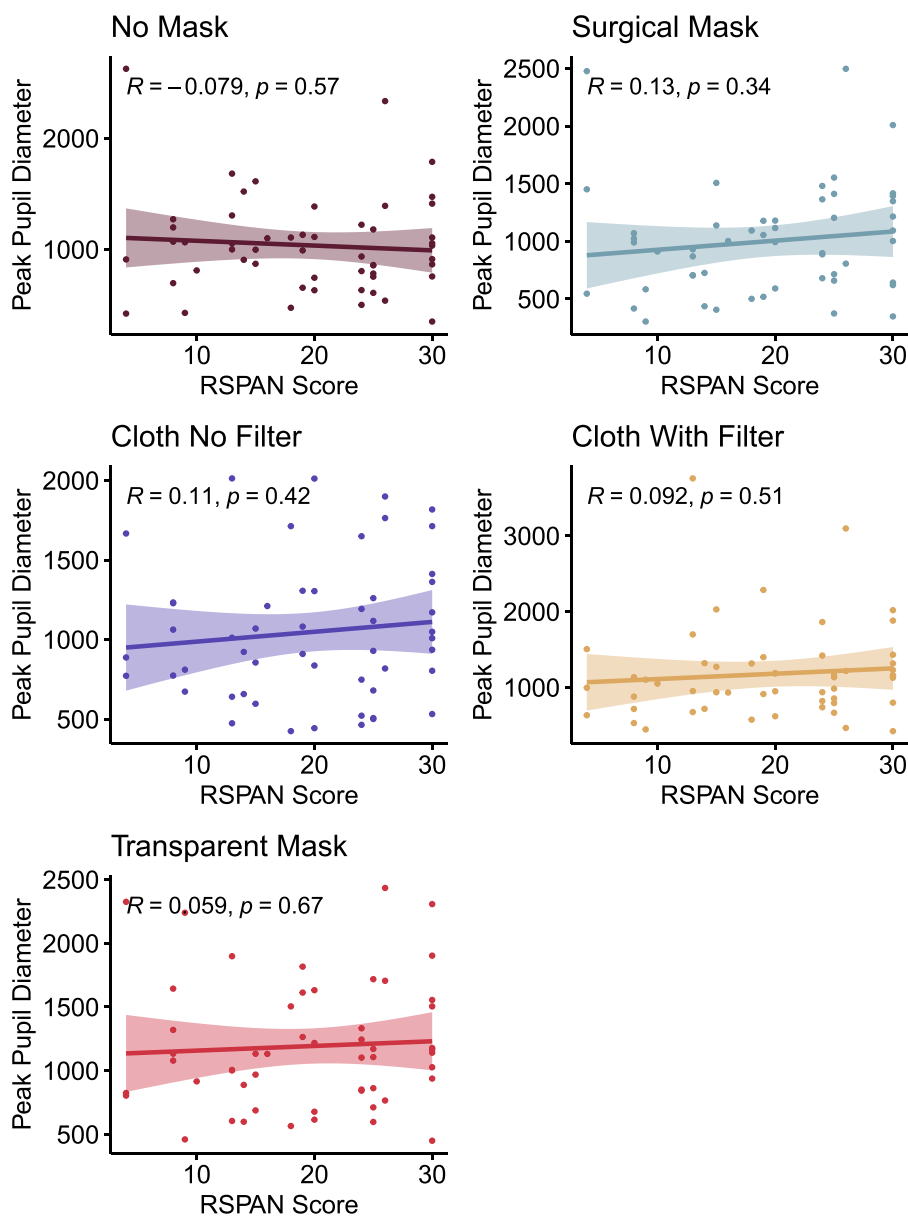


FIG. 7. (Color online) Correlations between Reading Span (RSPAN) scores and Peak Pupil Diameter in each mask condition.

pupil 500 times per second (down-sampled to 50 Hz for analyses) using a standard measure. As a result, the pupillometry data in the present study may provide a more nuanced depiction of cognitive effort. Moreover, pupil dilation is a non-volitional physiological response, whereas the subjective ratings are participants' reports of their conscious awareness of their ideas of "effort," and not all participants used the 1–100 scale in the same way. These differences in power and standard of measure could explain why the two measures did not correlate. Yet, despite these disparities, the two measures indicated the same general trend. Future studies might include both measures within subjects to demonstrate the extent to which subjective measures can be employed to unveil general trends. There are other summary statistics of pupil data that could be used to test for correlations between the pupil response and subjective ratings [e.g., mean pupil diameter, peak latency; see Zekveld *et al.* (2010)]. Future investigators may consider using these other

options. In our study, we chose peak pupil dilation and opted not to conduct additional analyses after seeing the results because of the increased likelihood of a type I error.

It is often considered a limitation of pupillometry that the task-evoked pupillary response is small relative to the pupil's response to light, because visual signals in combination with pupillometry pose challenges to experimental design. Indeed, the current study deviates from others on the topic of face masks in that it uses audio-only stimuli. One advantage to this design is that it greatly minimized external influences (such as listener expectations based on speaker appearance) and, therefore, allowed us to investigate more directly how the cognitive demands of processing speech filtered through face masks are driven by the acoustic signal itself. Indeed, Brown *et al.* (2021) provided an acoustic analysis of the long-term spectra of the stimuli used (replicated in this paper in Fig. 2), which showed the extent to which each mask muffled speech frequencies. Our

pupillometry data mirror those acoustic data: the mask that muffled speech frequencies the most was the most difficult to understand. Furthermore, the pupil data as portrayed in Fig. 3(a) (as well as the subjective effort and intelligibility data in Fig. 4) yield two noticeable “groupings” of the conditions marked by the division between the Cloth without Filter and Cloth with Filter conditions. The acoustic data do not show as clear a delineation, but beginning at ~3000 Hz, these same two conditions separate and remain separated (whereas other conditions overlap). The parity in the patterns across these sets of data support the notion that acoustics have a large effect on the cognitive demands of speech perception.

Last, it warrants discussing that in our data the transparent mask condition was associated with the most cognitive effort. This is not meant to negate the fact that listeners do benefit from the availability of a visual signal afforded by these types of masks (Atcherson *et al.*, 2017). Indeed, in their audiovisual study, Brown *et al.* (2021) report no differences in effort ratings between the Transparent Mask and Cloth Mask with Filter conditions, but our participants rated the audio-only Transparent Mask condition as significantly more effortful than the others. Finally, we emphasize that only one transparent mask was used in this study; materials and construction from other manufacturers may elicit different results.

VI. CONCLUSION

All face masks introduce some challenge to speech perception, but different masks impose different levels of impediment. We found that, of the different types of face masks we used and further masked in noise, the cognitive demands of listening to audio-only speech increased significantly in the following order: no mask < surgical mask < cloth mask without filter < cloth mask with filter < transparent mask.

Our results provide objective measures of the cognitive demands of listening to audio-only face mask-attenuated speech in noise, and expand upon previous subjective measures. Importantly, the current study focused on the acoustic signal itself and demonstrated that acoustics are a driving factor of the cognitive demands of degraded-speech perception when listening to a talker with a face mask.

The data that support the findings of this study are openly available on OSF (Carraturo, 2023).

ACKNOWLEDGMENTS

This research is partially supported by the Basque Government through the BERC 2022–2025 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S. There are no conflicts of interest to disclose. All procedures were approved by the Washington University in St. Louis Institutional Review Board and informed consent was obtained from all human subjects prior to participation.

APPENDIX

Stimuli List (originally from Van Engen *et al.*, 2012)

1. The hot sun warmed the ground
2. The gray mouse ate the cheese
3. The strong father carried my brother
4. The large monkey chased the child
5. The mean bear ate the fruit
6. The loud noise upset the baby
7. The friendly neighbor helped the grandmother
8. The black bear scared the visitors
9. The hungry children ate the snacks
10. The strong sister won the game
11. The rude joke upset my parents
12. The dark house scared the baby
13. The talented musician knew the songs
14. The gray horse ate the grass
15. The sick student read the book
16. The hungry girl made the sandwich
17. The tiny flies bothered the girl
18. The new student liked the professor
19. The hot coffee hurt the boy
20. The small animal scared the baby
21. The kind girl helped the strangers
22. the talented author received the prize
23. The black cat climbed the tree
24. The thoughtful boyfriend bought the flowers
25. The hungry dog ate the food
26. The friendly cat loved the boy
27. The old man cooked the carrots
28. The happy dog found the toy
29. The youngest sister watched the parade
30. The sweet dog watched the children
31. The pretty girl won the prize
32. The lonely artist called her friend
33. The youngest child hated the fruit
34. The cheap food attracted the customers
35. The rich boyfriend owned the houses
36. The new kitten climbed the tree
37. The angry bear scared the couple
38. The thirsty cat drank the milk
39. The three sisters shared the clothes
40. The tiny rabbit chewed the grass
41. The grocery store sold the food
42. The dangerous snake bit the rabbit
43. The troubled son stole the money
44. The hungry animal chewed the plants
45. The busy farmer grew the potatoes
46. The strong wind cleaned the air
47. The gray rabbit loved the carrots
48. The kind neighbor opened the door
49. The old garbage attracted the flies
50. The small boy chose the game
51. The large family expected the visitors
52. The small family played the game
53. The three sisters watched the movie
54. The tiny kitten chased the mouse

55. The angry husband visited the lawyer
56. The troubled child needed her mother
57. The pretty woman liked the cookies
58. The stormy weather destroyed the home
59. The interesting book saved the author
60. The sick patient received the flowers
61. The small restaurant needed the money
62. The talented artist made a picture
63. The English student read the book
64. The hot sun warmed the tea
65. The hardworking nurse helped the patient
66. The busy daughter joined the museum
67. The kind woman helped her neighbor
68. The talented doctor saved the child
69. The football player won the prize
70. The friendly waiters served the meal
71. The hardworking farmer grew the corn
72. The friendly girl shared the ball
73. The hungry rabbit ate the carrots
74. The friendly baby hugged the kitten
75. The helpful daughter cleaned the house
76. The thirsty dog drank the water
77. The rich man bought the wine
78. The soft music pleased the boss
79. The horrible story upset my grandmother
80. The mean children broke the rules
81. The teacher chose the horrible book
82. The children enjoyed the holiday parade
83. The girl loved the sweet coffee
84. The grandmother baked a sweet cake
85. The woman met the rich actor
86. The doctor owned the yellow car
87. The teacher wrote a difficult question
88. The store sold the dirty clothes
89. The ball broke the glass window
90. The grandfather loved the red wine
91. The brother met the talented artist
92. The chef baked the sweet corn
93. The father hugged his sad daughter
94. The chef cooked the delicious food
95. The bird found the juicy worm
96. The grandfather drank the dark coffee
97. The neighbor liked the loud song
98. The cat chased the gray mouse
99. The mother baked the delicious cookies
100. The team played a difficult game
101. The wind destroyed the tiny house
102. The restaurant sold the red wine
103. The musician played a beautiful song
104. The boy carried the heavy chair
105. The chef chose the delicious cheese
106. The man ate the large meal
107. The parents told the horrible story
108. The man shared the difficult story
109. The chef made the fresh noodles
110. The teacher read an interesting novel
111. The restaurant served a delicious soup
112. The woman heard a beautiful song
113. The grandmother loved the rich cake
114. The nurse cleaned the dirty clothes
115. The family watched the talented performer
116. The author told an interesting story
117. The painter owned the soft brushes
118. The store sold the delicious food
119. The travelers visited the new museum
120. The bird bothered the old dog
121. The customers hated the black tea
122. The mother drank the orange juice
123. The professor gave the unfair grade
124. The doctor helped the sick patient
125. The girl wanted the pretty flowers
126. The goat ate the sweet grass
127. The grandfather chose an old movie
128. The children liked the fresh vegetables
129. The boss met the new customers
130. The store sold the cheap picture
131. The monkey made the horrible noise
132. The artist visited the old museum
133. The dog chased the three rabbits
134. The man ate the fresh peppers
135. The boss told a horrible joke
136. The couple expected a new baby
137. The author wrote a long novel
138. The fans watched the football game
139. The boy carried the small rabbit
140. The mother bought the birthday gift
141. The family cleaned the dirty house
142. The dog bit the youngest boy
143. The son loved the toy car
144. The artist made a beautiful picture
145. The monkey wanted the yellow banana
146. The farmer grew the colorful peppers
147. The boss bought the new car
148. The school needed a new teacher
149. The president gave an interesting speech
150. The man heard the beautiful music
151. The mouse found the yellow cheese
152. The horse made the loud noise
153. The cat drank the fresh milk
154. The neighbor told an interesting story
155. The brother joined the soccer game
156. The player carried the soccer ball

¹See supplementary material at <https://doi.org/10.1121/10.0023953> for the plotted curves for additional plots and analyses.

²In Fig. 3(b), the data show differences in pupil response curves as a function of the intelligibility of each trial. The data show that low intelligibility trials yield later pupil dilation peaks and greater overall effort than high intelligibility. Because low-intelligibility trials represent a small subset of all trials, the mean curves [Fig. 3(a)] peak and fall sooner [i.e., matching the high-intelligibility data in Fig. 3(b)].

Atcherson, S. R., Mendel, L. L., Baltimore, W. J., Patro, C., Lee, S., Pousson, M., and Spann, M. J. (2017). "The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss," *J. Am. Acad. Audiol.* **28**(1), 058–067.

- Audacity Team (2021). "Audacity(R): Free audio editor and recorder [computer application]," version 3.0.0, <https://audacityteam.org/> (Last viewed March 17, 2021).
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Soft.* **67**(1), 1–48.
- Borrie, S. A., Barrett, T. S., and Yoho, S. E. (2019). "Autoscore: An open-source automated tool for scoring listener perception of speech," *J. Acoust. Soc. Am.* **145**(1), 392–399.
- Brown, V. A., Van Engen, K., and Peelle, J. E. (2021). "Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults," *Cogn. Res.* **6**(1), 1–12.
- Carraturo, S. (2023). "Face mask attenuated speech & listening effort," <https://osf.io/z8gt5/> (Last viewed 12/15/2023).
- Carraturo, S., McLaughlin, D. J., Peelle, J. E., and Van Engen, K. (2023). "Face mask attenuated speech & listening effort," <https://osf.io/egp8d> (Last viewed 12/15/2023).
- Corey, R. M., Jones, U., and Singer, A. C. (2020). "Acoustic effects of medical, cloth, and transparent face masks on speech signals," *J. Acoust. Soc. Am.* **148**(4), 2371–2375.
- Geller, J., Winn, M. B., Mahr, T., and Mirman, D. (2020). "GazeR: A package for processing gaze position and pupil size data," *Behav. Res.* **52**, 2232–2255.
- Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., and Pavani, F. (2021). "Unmasking the difficulty of listening to talkers with masks: Lessons from the COVID-19 pandemic," *i-Perception* **12**(2), 204166952199839.
- Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Adv. Psychol.* **52**, 139–183.
- Kahneman, D., and Beatty, J. (1966). "Pupil diameter and load on memory," *Science* **154**(3756), 1583–1585.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). "Pupil dilation uncovers extra listening effort in the presence of a single-talker masker," *Ear Hear.* **33**(2), 291–300.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). "lmerTest Package: Tests in linear mixed effects models," *J. Stat. Soft.* **82**(13), 1–26.
- Lee, E., Cormier, K., and Sharma, A. (2022). "Face mask use in healthcare settings: Effects on communication, cognition, listening effort and strategies for amelioration," *Cogn. Res.* **7**(1), 1–9.
- Magee, M., Lewis, C., Noffs, G., Reece, H., Chan, J. C., Zaga, C. J., Paynter, C., Birchall, O., Azocar, S. R., Ediriweera, A., Kenyon, K., Caverlé, M. W., Schultz, B. G., and Vogel, A. P. (2020). "Effects of face masks on acoustic analysis and speech perception: Implications for peripandemic protocols," *J. Acoust. Soc. Am.* **148**(6), 3562–3568.
- Mathôt, S. (2018). "Pupillometry: Psychology, physiology, and function," *J. Cogn.* **1**(1), 1–23.
- McLaughlin, D. J., and Van Engen, K. J. (2020). "Task-evoked pupil response for accurately recognized accented speech," *J. Acoust. Soc. Am.* **147**(2), EL151–EL156.
- McLaughlin, D. J., Zink, M. E., Gaunt, L., Spehar, B., Van Engen, K. J., Sommers, M. S., and Peelle, J. E. (2022). "Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults," *Psychon. Bull. Rev.* **29**(1), 268–280.
- McLaughlin, D. J., Zink, M. E., Gaunt, L., Reilly, J., Sommers, M. S., Van Engen, K. J., and Peelle, J. E. (2023). "Give me a break! Unavoidable fatigue effects in cognitive pupillometry," *Psychophysiology* **60**, e14256.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R* (CRC Press, New York).
- Nichols, A. L., and Maner, J. K. (2008). "The good-subject effect: Investigating participant demand characteristics," *J. Gen. Psychol.* **135**(2), 151–166.
- Oswald, F. L., McAbee, S. T., Redick, T. S., and Hambrick, D. Z. (2015). "The development of a short domain-general measure of working memory capacity," *Behav. Res.* **47**(4), 1343–1355.
- Peelle, J. E. (2018). "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior," *Ear Hear.* **39**(2), 204–214.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., 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., and Wingfield, A. (2016). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," *Ear Hear.* **37**, 5S–27S.
- Psychology Software Tools, Inc. (2012). [E-Prime 2.0], <https://support.pst-net.com/> (Last viewed 12/15/2023).
- Reilly, J., Kelly, A., Kim, S. H., Jett, S., and Zuckerman, B. (2019). "The human task-evoked pupillary response function is linear: Implications for baseline response scaling in pupillometry," *Behav. Res.* **51**, 865–878.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (2013). "The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances," *Front. Syst. Neurosci.* **7**(31), 1–17.
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC.
- Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., and Smith, J. (2018). "Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures," *J. Speech. Lang. Hear. Res.* **61**(6), 1463–1486.
- Van Engen, K. J., and McLaughlin, D. J. (2018). "Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition," *Hear. Res.* **369**, 56–66.
- Van Engen, K. J., Chandrasekaran, B., and Smiljanic, R. (2012). "Effects of speech clarity on recognition memory for spoken sentences," *PLoS One* **7**(9), e43753.
- Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). "The impact of auditory spectral resolution on listening effort revealed by pupil dilation," *Ear Hear.* **36**(4), e153.
- Winn, M. B., and Teece, K. H. (2021). "Listening effort is not the same as speech intelligibility score," *Trends Hear.* **25**, 233121652110276.
- Yi, H., Pingsterhaus, A., and Song, W. (2021). "Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic," *Front. Psychol.* **12**, 682677.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). "Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response," *Ear Hear.* **32**(4), 498–510.
- Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). "The pupil dilation response to auditory stimuli: Current state of knowledge," *Trends Hear.* **22**, 233121651877717–233121651877766.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Ear Hear.* **31**(4), 480–490.