Volume 45 http://acousticalsociety.org/

181st Meeting of the Acoustical Society of America

Seattle, Washington

29 November - 3 December 2021

Psychological and Physiological Acoustics: Paper 2pPPb10

Listening effort elicited by energetic versus informational masking

Sarah Villard and Tyler Perrachione
Department of Speech, Language, & Hearing Sciences, Boston University, Boston, MA, 02215; sarah.n.villard@gmail.com; tkp@bu.edu

Sung-Joo Lim
Department of Psychology, Binghamton University, Binghamton, NY; sungjoo@binghamton.edu

Ayesha Alam and Gerald Kidd
Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, 02215; ayeshalm@bu.edu, gkidd@bu.edu

Measuring the *listening effort* exerted by an individual to understand speech under auditory masking conditions can provide vital information not available from speech intelligibility scores alone. The goal of this study was to compare the amount of listening effort elicited in young, normal-hearing subjects during an intelligible speech masking condition producing a high degree of informational masking (due to listener uncertainty, e.g. target-masker confusions) versus in two speech spectrum-shaped noise masking conditions producing primarily energetic masking (due to spectrotemporal overlap between target and maskers). One noise masking condition involved unmodulated noise, while the other involved speech-envelope-modulated noise. Listening effort was measured simultaneously using two physiological metrics: pupil dilation and alpha power (via electroencephalography). The target speech comprised 5-word matrix-style sentences in both conditions. In each condition, the target-to-masker ratio was set at each subject's 75% correct point on the psychometric function. Target and maskers were spatially separated. Results indicated that the intelligible speech masking condition elicited a greater change in pupil size than the unmodulated noise masking condition. This finding is consistent with the view that greater effort is involved in ignoring acoustically and linguistically similar sources than highly dissimilar, low-information value sources. Work supported by: NIH K99DC018829.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001546

INTRODUCTION

A growing body of research has demonstrated that the degree of cognitive resources deployed by a listener when attending to and processing target speech, termed "listening effort," is dissociable from speech intelligibility (Koelewijn et al, 2012; Winn & Teece, 2021). Assessing listening effort—especially under adverse conditions—may therefore provide valuable information about the experience of the listener that is not available when the sole measure of performance is speech intelligibility (Rennies et al., 2014; 2019). Quantifying the amount of listening effort expended during a given task allows us to move beyond the standard question, *Under what conditions can the listener solve this task?* to ask, *What does solving this task demand of the listener?* Addressing this second question adds a new dimension to the study of auditory masking, degradation, and other challenging manipulations of speech stimuli.

Measuring listening effort has the potential to shed light on hidden psychological variability underlying successful speech perception outcomes. However, quantitative and objective measures to operationalize this construct have remained elusive. Measures of listening effort currently in use include behavioral self-report, as well as several physiological indices including pupillometry (which measures change in pupil size) and electroencephalography (EEG, which measures neural oscillatory power). Interestingly, studies assessing listening effort using multiple indices have generally found a lack of correlation between them (e.g., Alhanbali et al, 2019; McMahon et al, 2016; Miles et al, 2017; Visentin et al, 2021), indicating that listening effort may be a multifaceted construct that cannot be characterized through a single behavioral or physiological approach (Alhanbali et al, 2019; Francis & Love, 2020). Studies examining listening effort using more than one approach have the potential to shed additional light on the possible relationships between different aspects of this complex phenomenon. In the current study, pupillometry and EEG recordings were obtained simultaneously in order to produce two separate sets of measurements that could readily be compared to one another.

Gaining a better understanding of listening effort may be particularly useful in situations where speech is masked by competing sounds. A great deal of typical human conversation takes place in the presence of extraneous background voices and/or "noise", which must be filtered out by the listener, a dilemma often referred to as the “cocktail party problem” (Cherry, 1953). Determining how listening effort fits into the cocktail party problem may lead to a better understanding of when and why breakdowns in processing and comprehension in non-ideal listening situations occur, as well as how to mitigate them. Auditory masking typically is considered to comprise both energetic masking (EM, or the result of spectrotemporal overlap between target and masker energy) and informational masking (IM, or additional masking that cannot be explained by spectrotemporal overlap), which usually are thought to arise predominantly from peripheral and central mechanisms within the auditory system, respectively (cf. Kidd et al., 2008). Most masking conditions include some degree of both EM and IM. However, it is generally the case that noise-masking conditions (referring to Gaussian noise) produce primarily EM, whereas speech masking conditions—particularly when the masking speech is intelligible—produce higher amounts of IM. The IM that is present when speech masks other speech results in part from confusions between target and masker sources on the part of the listener, despite adequate audibility of the target (see Kidd & Colburn, 2017, for a review of IM under speech masking conditions). In examining the effect of auditory masking on listening effort, a central question is whether (and if so, how) the listening effort needed to overcome EM vs. IM depends on different underlying processes. One previous study has provided strong evidence that IM is associated with increased self-reported listening effort (Rennies et al, 2019); the current study aimed to further investigate this topic using physiological measures.

A number of previous studies using physiological indices of listening effort have provided some insight into the effects of EM vs. IM by including both noise masking and speech masking conditions (e.g., Koelewijn et al, 2012; Versfeld et al, 2021; Wendt et al, 2018). However, none of these studies used speech masking conditions that both (a) maximized the amount of IM in the signal and (b) directly compared the effects of speech vs. noise maskers. Several studies used speech maskers that consisted of, for example, a randomly-chosen portion of a string of concatenated sentences (Koelewijn et al, 2012; Versfeld et al, 2021), or one or more talkers reading the newspaper (Wendt et al, 2018). While such maskers surely produced some IM, their differences from the short target sentences with respect to multiple source segregation cues (such as differences in intonation, syntactic structure, topic, rate of speech, etc.), likely made these stimuli comparatively easy to distinguish on the majority of trials (cf. Mattys et al.,

2012; Kidd and Colburn, 2017). Several other listening effort studies have used intelligible speech (or unintelligible "babble") maskers but not noise maskers, limiting any direct comparison between the effects of EM and IM (e.g., Bönitz et al, 2021; Lau et al, 2019). At least two other physiological listening effort studies have used vocoded target speech along with vocoded maskers (McMahon et al, 2016; Miles et al, 2017), which likely created some amount of confusability between similar-sounding target and maskers, and hence IM. However, the use of vocoding complicates the estimation of the amount of IM because it also affects the amount of EM presumably due to increased overlap between channels, and, furthermore, no stationary noise maskers were used for comparison. One previous study did involve a condition where a single, complete, intelligible masker sentence was presented simultaneously with a single complete target sentence (Kerlin, Shahin, & Miller, 2010); however, in that study the target and masker sentences always began with two consistent and distinct carrier phrases, likely increasing their distinguishability. In short, we are not aware of any study that has compared the effects of EM and IM on physiological indices of listening effort using a speech condition involving highly confusable target and masker sentences and incorporating a spectrotemporally similar noise masking condition for comparison. The primary goal of the current study, therefore, was to address this gap in the literature by measuring listening effort under carefully-constructed high-EM and high-IM masking conditions. A related secondary area of inquiry was whether different indicators of listening effort would be differentially sensitive to high EM vs. high IM conditions.

When designing a listening effort study, it is essential to take into consideration that the expenditure of effort is under volitional, top-down control by the listener. Dimitrijevic and colleagues (2019) found that when listeners were presented with digits in the presence of background noise and told to listen to and report the digits, listening effort (as indexed by a change in alpha power using EEG) increased; however, when they were presented with the same auditory stimuli and told to watch a silent movie instead of paying attention, no change in this index of listening effort was observed. In a different study, Zhang and colleagues (2019) found that offering listeners a monetary reward for better performance resulted in greater increases in listening effort (as indexed in this case by changes in pupil size). Additionally, Rowland and colleagues (2018) found notable prefrontal activation during a selective listening task, indicating the involvement of higher-level cognitive processes. The decision to voluntarily expend effort has also been shown to be sensitive to task difficulty: listening effort first increases as a simple task becomes more challenging, but may then begin to decrease again after the level of difficulty exceeds a certain point (Lau et al., 2019; Ohlenforst et al, 2018; Wendt et al, 2018), likely because listeners become frustrated and "give up," resulting in the exertion of less effort (Wendt et al, 2018; Winn et al 2018). It is therefore important to ensure that a task intended to elicit measurable listening effort is neither so easy that it does not require much effort nor so difficult that it causes listeners to become overly frustrated and disengage.

A standard approach to controlling task difficulty in auditory masking studies is to adjust the target-to-masker ratio (TMR) when measuring listening effort. Several studies on listening effort have used different fixed TMRs across participants; for example, Lau and colleagues (2019) presented target and maskers at +6 dB TMR and 0 dB TMR to all participants, and Wendt et al (2018) used a series of different fixed ratios between -20 dB and +8 dB with the aim of mapping a psychometric function of effort. A larger group of studies, however, have used individually-estimated TMRs for particular points on the psychometric function for individual listeners, as a means of demonstrating that listening effort may vary across participants and conditions even when performance levels are held constant. Typically, individually-estimated TMRs corresponding to points on the psychometric function between 50% and 90% correct have been selected (e.g., Alhanbali et al, 2019; Bönitz et al, 2021; Dimitrijevic et al, 2019; McMahon et al, 2016; Miles et al, 2017; Wendt et al, 2018). Ohlenforst and colleagues (2018) observed peak pupil dilations at approximately 50% correct TMRs for sentence recognition, regardless of masking condition. The current study used a similar approach, presenting participants with TMRs corresponding to their individually-estimated 75% correct point for each condition.

The goal of the current study was to build on the existing research on listening effort in the presence of auditory masking by using concurrent pupillometry and EEG to measure and compare the amount of listening effort exerted in carefully-controlled high-IM versus high-EM listening conditions. Findings from previous work regarding optimal calibration of task difficulty were applied in order to keep listeners engaged but not frustrated. Specifically, the study had three aims:

1. To measure the effect of type of masking (EM vs. IM) on change in alpha power, as measured by EEG. It was hypothesized that a high-IM condition would result in greater changes in alpha power, relative to a high-EM condition.
2. To measure the effect of type of masking (EM vs. IM) on change in pupil size, as measured by pupillometry. It was hypothesized that a high-IM condition would result in greater changes in pupil size, relative to a high-EM condition.
3. To determine whether there is a relationship between changes in pupil size and alpha power under high-IM versus high-EM conditions. Based on previous work comparing the results of co-registered pupillometry and EEG, no significant correlation between change in pupil size and change in alpha power was expected.

METHODS

A. PARTICIPANTS

Fifteen young adult, normal-hearing listeners participated in this experiment (mean age = 20.8, range = 18-24, 5M 10F). Participants were recruited through the existing database of the Psychoacoustics Laboratory at Boston University, as well as through online postings at Boston University. All participants demonstrated normal hearing, defined as thresholds of 20 dB HL or better at 0.25, 0.5, 1, 2, 4, and 8 kHz, in both the left and right ears, during a pure tone audiometric screening. All participants were native English speakers who reported no diagnosis of attention deficit disorder or history of head injury resulting in loss of consciousness. This study was approved and overseen by the Boston University Institutional Review Board.

B. EXPERIMENTAL STIMULI AND CONDITIONS

Auditory stimuli in the current experiment (see Table 1) consisted of a set of recordings of one-syllable words drawn from a corpus that has been used in a number of previous studies in our laboratory (e.g., Kidd et al, 2016), as well as some additional recordings of two-syllable names drawn from a separate corpus but spoken by the same talkers. Target stimuli consisted of single-word recordings concatenated into 5-word sentences. The target sentence for each trial always began with the designated target cue word “Sue”, followed by a verb, number, adjective, and noun, with each of the latter 4 words randomly selected from the list of 8 possibilities. Examples of possible target sentences included “Sue sold eight red toys” or “Sue found three old gloves”. Each target sentence was spoken by a single female talker, randomly selected on each trial from a pool of 8 female talkers.

Names	Verbs	Numbers	Adjectives	Objects	Additional 2-syllable names
Bob	bought	two	big	bags	Allen
Jane	found	three	cheap	cards	Doris
Jill	gave	four	green	gloves	Kathy
Lynn	held	five	hot	hats	Lucy
Mike	lost	six	new	pens	Peter
Pat	saw	eight	old	shoes	Rachel
Sam	sold	nine	red	socks	Thomas
Sue	took	ten	small	toys	William

Table 1: List of all words used in the experiment

Each trial also included the presentation of two masker strings. Depending on the listening condition, these maskers either consisted of intelligible speech (IntSpeech), speech-spectrum-shaped, speech-shaped stationary (i.e., unmodulated) noise (StatNoise), or speech-spectrum-shaped, speech-envelope-modulated noise (ModNoise). The

rationale for using two different high-EM noise conditions was that the StatNoise condition was similar to the continuous noise masking conditions used in a number of previous studies on listening effort (e.g., Alhanbali et al, 2019; Dimitrijevic et al, 2019) and so would yield results that would be directly comparable to results of those studies, while the ModNoise condition contained the same broadband temporal envelopes (i.e., single-channel vocoder) as the speech masking condition used in the current study. Visual stimuli consisted of columns of typed words presented as response options, arranged vertically in a GUI. All stimuli were presented using custom software in MATLAB (MathWorks, Inc., Natick, MA) and Psychtoolbox-3.

C. PROCEDURES

Throughout all portions of the experiment, participants were seated in a dimly-lit, sound-attenuated booth, in front of a computer monitor. The stimuli were presented through three loudspeakers, each located approximately 1.5 meters from the listener's head and positioned at 0° and ±45° degrees azimuth. Target stimuli were always presented from the loudspeaker located at 0° azimuth. Maskers, when present, were always presented simultaneously from the two loudspeakers located at ±45° degrees azimuth (see Figure 1a). The experiment consisted of behavioral testing, followed by a pupillometry/EEG recording session.

Across all conditions, onset of the two maskers always preceded onset of the target sentence (see Figure 1b for a depiction of the trial structure). In the IntSpeech condition, each masker consisted of an 8-word string, beginning with 3 randomly-drawn 2-syllable names, followed by a 5-word sentence drawn from the same matrix as the target sentences. A sample masker string might be "William Peter Kathy Bob sold three red toys", with the latter 5 words presented simultaneously with the 5-word target sentence, with the onsets of each target and masker word aligned. Within a given trial, the words chosen for Masker1, Masker2, and the target sentence were selected randomly without replacement and thus were mutually exclusive; similarly, the Masker1 talker, Masker2 talker, and target talker also were mutually exclusive random selections within a given trial.

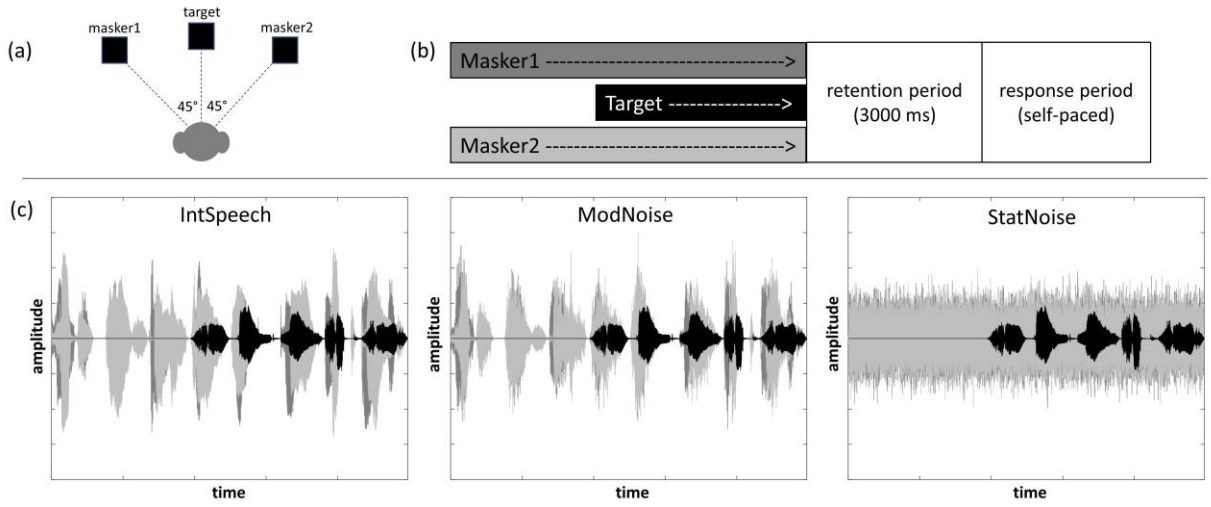

Figure 1a) Loudspeaker setup; Figure 1b) Trial structure across all conditions; Figure 1c) Waveforms of audio stimuli for each condition (10 dB TMR); target in black.

In the StatNoise condition, each string of eight masker words was replaced by a single continuous token of stationary noise with a spectrum matching that of the speech corpus. In the ModNoise condition, each masker consisted of this same token of stationary noise, modulated using the envelopes of eight masker words (e.g., the envelopes of "William Peter Kathy Bob sold three red toys" spoken by a randomly-selected talker). The onset of the target relative to the onset of the maskers varied based on the specific 2-syllable names chosen during a given trial,

but the average target onset time was 2115 ms (sd: 109 ms) after the onset of the maskers. Following a 3000 ms[1] retention period, participants were presented with a series of 5 response GUIs following presentation of the auditory stimulus, one for each word in the target sentence. Each response GUI contained all eight possibilities for that word (see Table 1), with the exception of the cue word, where "Sue" was the only option provided. Participants used a mouse click to select their response from each GUI. See Figure 1c for sample waveforms for each condition.

Behavioral testing. During the behavioral portion of the experimental task, participants first completed a quiet (masker-free) practice block consisting of 10 trials to familiarize them with the target sentences used in the experiment. Next, participants completed two 1-up, 1-down quiet adaptive tracks designed to estimate speech reception thresholds (SRTs) for these sentences. Throughout the behavioral experiment, participants were instructed to maintain the same seating position to ensure consistent spatial presentation of stimuli relative to the head.

After completing these preliminary blocks, participants completed 3 experimental blocks in each of the 3 masking conditions, with block order counterbalanced across participants. In an adaptation of Brand and Kollmeier's (2002) method, each 30-trial block consisted of two separate but randomly interleaved 15-trial tracks designed to estimate the TMRs corresponding to two different points on the psychometric function: the TMR at which the participant was predicted to achieve 75% correct intelligibility and the TMR at which the participant was predicted to achieve 25% correct intelligibility. Note that the first word, *Sue,* was always given and therefore never scored. Target sentences were always played at 40 dB SPL; masker sentence levels were varied to achieve the required TMR for each trial. At the end of each block, the last three TMRs of each track were averaged to produce a 75% correct and a 25% correct estimate. The 3 estimates for each condition were then averaged to produce one overall 75% correct estimate and one overall 25% correct estimate for each participant in each condition.

EEG/pupillometry recording session. All participants completed the entirety of the simultaneous EEG/pupillometry recording during a single study visit, which occurred on a separate day following the completion of their behavioral testing sessions. During the EEG-pupillometry session, participants completed two blocks in each condition[2], with block order counterbalanced across participants. The TMR used for each condition was equal to the participant's overall 75% correct estimate in that condition and was kept constant throughout the block. Target sentences were always played at 40 dB SPL. The retention period was 3000 ms for all participants.

A Biosemi ActiveTwo system was used to acquire EEG data (sampling rate: 2048 Hz), using 32 scalp channels configured in the standard 10/20 montage, as well as 4 additional external facial electrodes placed to record eye movements and 2 electrodes placed on the mastoids for reference. Additionally, an SR Research Eyelink 1000 was used during this session to record changes in pupil diameter (sampling rate: 500 Hz). A stationary chinrest was used to stabilize head position. Prior to the start of each block, the participant completed an eye gaze calibration. Prior to each trial, a drift correction check was performed to ensure that the participant's gaze was directed toward the center of the screen. Following the drift check, the participant clicked the mouse to start the trial. From this point onward, the same trial structure as the behavioral sessions was used (see Figure 1b). Participants were instructed to look at the center of the screen whenever possible, with the exception of the response periods. Because the EEG data for all 6 blocks were acquired as a single, continuous file, participants were asked to remain seated throughout all blocks; however, they were encouraged to take seated rest breaks between blocks as needed. In addition to the EEG and pupillometry data, behavioral accuracy data were collected during this session.

DATA ANALYSIS

EEG data analysis. EEG data were preprocessed using EEGLAB (Delorme & Makieg, 2004), Fieldtrip (Oostenveld et al, 2011), and customized MATLAB scripts. Initial preprocessing steps consisted of referencing to the external mastoid electrodes, resampling to 256 Hz, and bandpassing the data from 1 to 30 Hz. Next, an independent components analysis was performed for each participant to identify blinks and saccades. A blink

[1] Participant 1 completed the adaptive tracks with a retention period of 6000 ms. This period was deemed to be unnecessarily long and was shortened to 3000 ms for all subsequent participants.

[2] For Participant 8, all data from one of the ModNoise conditions failed to save. This participant returned to redo this block on a different day; however, for logistical reasons, only pupillometry data were able to be recorded during this redo session. Therefore Participant 8 only contributed one ModNoise block to the EEG data.

component was identified and removed for each participant, and a saccade component was identified and removed for 13 of the 15 participants. Next, any remaining noisy channels were removed from the data and interpolated based on neighboring channels. The continuous data were then epoched into individual trials, and trials where the voltage range exceeded 200 μV were excluded from further analysis. Next, a power computation was performed for each time-frequency bin (100 ms x 1 Hz) within each trial, within each channel. An average baseline across trials was calculated separately for each condition, based on data from the last 1000 ms of the masker-only, pre-target listening portion of the trials. Finally, a divisive baseline correction was performed, such that the corrected value in a given time-frequency bin represented the percent change from baseline.

Pupillometry data analysis. Pupil diameter measurements, in arbitrary units (AU) measured by the Eyelink 1000, were preprocessed using the R package GazeR (Geller et al, 2020). Only measurements for the right pupil were included in the analysis. Initial preprocessing steps for each participant included identification and extension of blinks (100 ms pre- and post-blink), interpolation of data during and surrounding blink periods, and smoothing of the pupil trace (using a 5-point moving average). Removal of outliers was then performed by visually inspecting a histogram of all raw pupil sizes for a given participant and determining upper and lower cutoff points, followed by a median absolute deviation analysis, which removes additional outliers by identifying rapid temporal changes in pupil size. Next, a subtractive baseline correction was performed for each trial. As with the EEG analysis (above), the last 1000 ms of the masker-only, pre-target listening portion of each trial was used as a baseline. For the pupillometry analysis, the median (rather than mean) baseline value was used, as recommended by Mathôt and colleagues (2018). Following baseline correction, pupil data samples were averaged into bins of 100 ms.

RESULTS

Behavioral data collected during EEG/pupillometry recording session. Accuracy data from the physiological recording session was calculated at the word level for each participant. The mean accuracy across participants was 72.2% for IntSpeech, 70.8% for ModNoise, and 73.2% for StatNoise.

EEG data. A heat map displaying baseline-corrected and time-binned percent change in alpha power (8-13 Hz) for each condition, using all available channels from each participant, and averaged across 14 participants, is presented in Figure 2a. To compare change in alpha power between the three conditions, a 1 x 3 repeated-measures analysis of variance (RM-ANOVA) was performed on the mean baseline-corrected alpha power for the region of interest between 0 ms and 4000 ms relative to target onset, within the frequency band range 8-13 Hz (this region is bounded by black rectangles in Figure 2a). Results of the RM-ANOVA were non-significant (see Figure 2b).

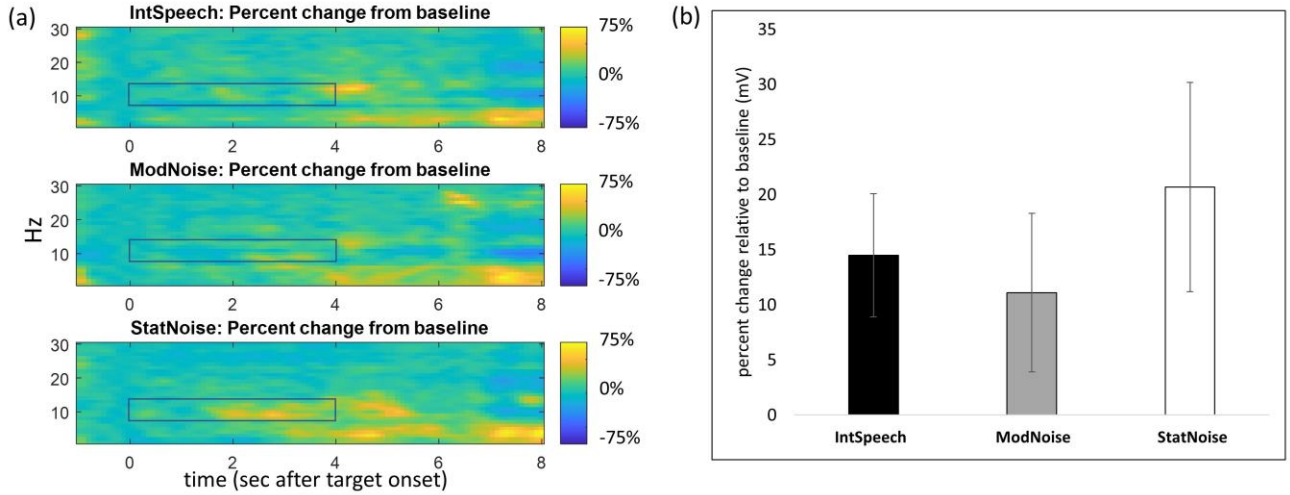

Figure 2a: Time-course of change in alpha power relative to baseline. Figure 2b: Mean baseline-corrected alpha power (error bars depict +/- 1 standard error.)

Pupillometry data. Figure 3 depicts the baseline-corrected and time-binned time courses of the pupil traces for each condition, averaged across all 15 participants. Using the existing pupil size literature as a guide, three outcome measures were selected from the baseline-corrected and binned data for statistical comparisons between conditions: (1) mean pupil size from 0 ms to 6000 ms relative to target onset, (2) peak pupil size during this period, and (3) latency to peak. Note that a longer time period of interest relative to target onset was used in this analysis relative to the EEG analysis (0-6000 ms rather than 0-4000 ms) because the pupil response is known to take longer to peak. For each outcome measure, a 1 x 3 RM-ANOVA was performed to compare the effect of masking condition on that measure (see Figure 4). The RM-ANOVA examining the effect of condition on peak pupil size produced a significant result, $F(2,28) = 5.26$, $p < 0.05$, with post-hoc pairwise testing revealing a significant difference only between IntSpeech and StatNoise ($p < 0.001$). Results of the other two RM-ANOVAs were not significant.

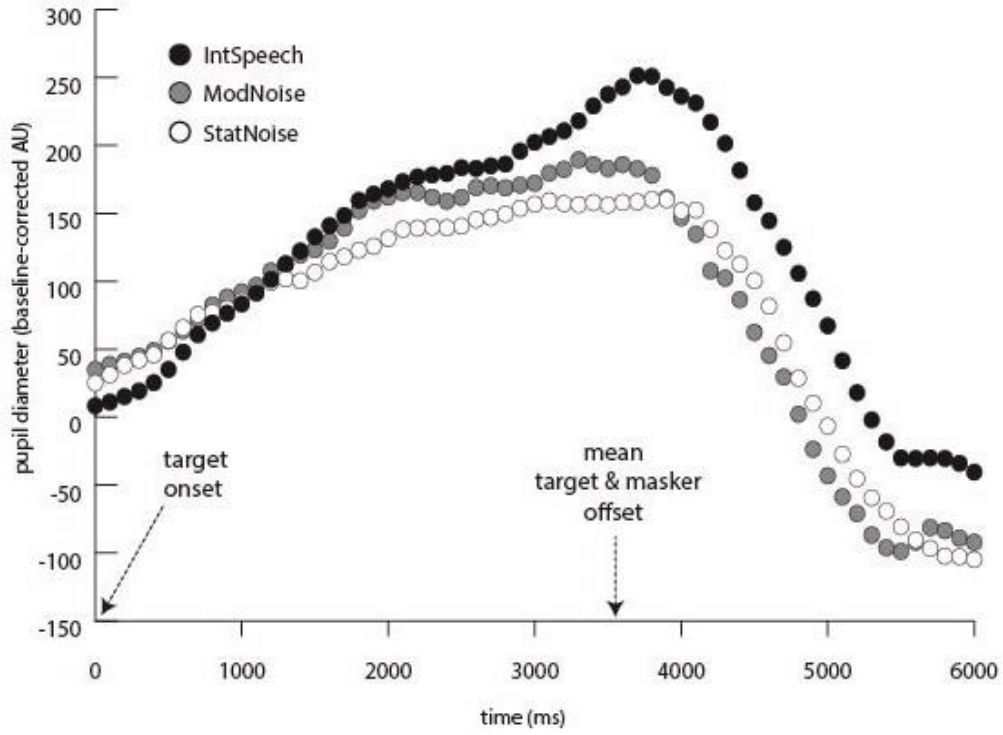

Figure 3: Baseline-corrected time courses of pupil traces, averaged across participants

Comparison of EEG and pupillometry results. In order to determine whether there was an association between the baseline-corrected mean alpha power values from 8-13 Hz (from 0-4 seconds following the onset of the target) and the baseline-corrected mean pupil size values (from 0-6 seconds following the onset of the target), three Pearson correlations were performed, one for each condition. None of these three correlations returned a significant result.

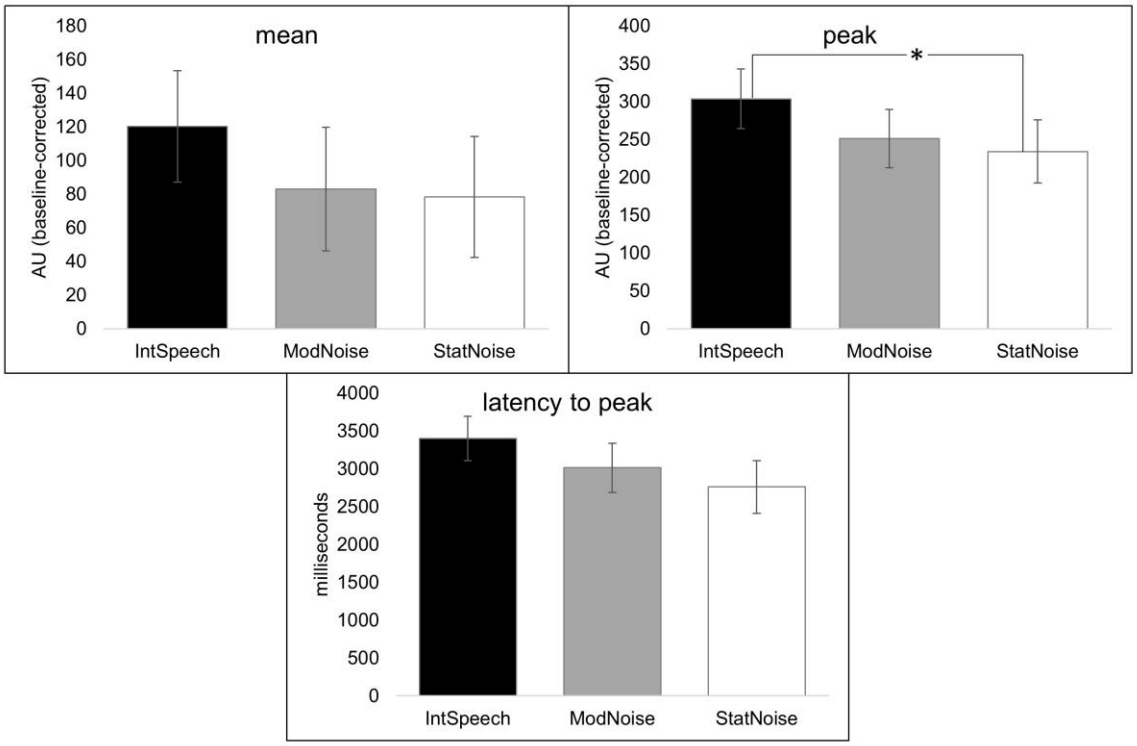

Figure 4: Mean, peak, and latency to peak for each condition (error bars depict +/- 1 standard error)

DISCUSSION & CONCLUSION

This study examined the effect of masker type on listening effort in young, normal-hearing individuals, with the goal of comparing the effect of a high-informational masking condition with the effects of two high-energetic masking conditions. The study measured listening effort using two frequently-employed physiological indices, alpha power (measured using EEG), and pupil size. Results of the study indicate that masking type does have a significant effect on listening effort as indexed by change in pupil size but not as indexed by alpha power. No significant relationship between measurements of listening effort obtained using these two different indices was observed.

The pupillometry findings regarding peak pupil size aligned with the study hypothesis that a high-IM condition would result in significantly greater changes in pupil size than a high-EM condition. This suggests that listening conditions involving highly confusable competing speech stimuli may require, or elicit, more effort on the part of the listener than a listening condition involving continuous unintelligible noise stimuli, at corresponding performance levels. Despite this finding, however, several important questions about these data remain. To begin with, it is not clear why post-hoc testing revealed a significant difference between the intelligible speech masking condition and the unmodulated noise masking condition, but not between the intelligible speech masking condition and the modulated noise masking condition. Since the latter difference was close to being significant, this could be due to the sample size of the study. Another possible explanation (not mutually exclusive with the first) is that the modulated noise maskers may have contained additional masking not present in the unmodulated noise masking condition (i.e., due to the modulation itself, see Conroy & Kidd, 2021; Stone & Moore, 2014). If this was indeed the case, it would be consistent with the conclusion that increased IM results in increased listening effort.

It is important to note that the statistical outcomes reported here were dependent in part on the baseline correction method that was selected and applied during preprocessing of pupillometric data. Visual inspection of shapes of the raw pupil traces in each condition, examined beginning with the onset of the maskers, suggested that the pupil traces in the intelligible speech masking condition differed notably from those in the noise masking conditions during the baseline period. (These raw pupil traces are not depicted in Figure 3, which starts at the target

onset and displays baseline-corrected measurements only.) Specifically, in the intelligible speech masking condition, the average raw pupil traces increased beginning with the onset of the maskers but then dipped back downward approximately 500 ms prior to the target onset, whereas in the noise masking conditions, the average raw pupil traces rose continuously throughout the masker-only listening period. The reason for the differences in raw pupil traces between conditions is unknown; additional data on masker-only pupil size would be needed in order to explain this phenomenon. However, one important takeaway from this examination of the raw data is that different types of auditory maskers may have different effects on pupil size. This in turn suggests that decisions made when selecting the length of the baseline (masker-only) period and method of baseline correction could affect the results.

In contrast to the pupillometry analysis, the EEG analysis did not reveal any significant difference between conditions. Examination of the variability in the data suggests that a larger sample size might be needed in order to detect a significant difference (if any exists). However, it is also worth noting that (like the pupil size analysis) the EEG analysis could reasonably have been conducted using a different baseline calculation. A different or more specific time-frequency region of interest could also have been selected.

The lack of association between pupil size data and EEG data aligned with the study hypothesis, as well as with previous studies that also have found no association between different indices of listening effort. As more and more studies report a lack of association, the possibility that the construct of "listening effort" in fact encompasses several different constructs. Interestingly, in the current study, not only was no significant correlation observed, but visual inspection of the EEG heat maps suggests that, if anything, the intelligible speech masking condition resulted in a smaller change in alpha power than either of the two noise masking conditions.

One of the major strengths of the current study was its use of carefully-controlled linguistic and non-linguistic stimuli, which allowed for a comparison between closely-aligned high-EM and high-IM listening conditions. On a related note, this study also employed spatial separation between target and maskers, which provides binaural spatial cues that listeners may use to perceptually separate target and maskers, resulting in a release from masking. This release from masking is particularly pronounced in speech-on-speech masking (high-IM) conditions (Arbogast et al, 2005). While a small number of studies on listening effort under auditory masking conditions have implemented spatial separation between target and masker (e.g., Kerlin et al, 2010; Wendt et al, 2018), other studies have presented stimuli either monaurally (e.g., Versfeld et al, 2021) or diotically (e.g., Dimitrijevic et al, 2019; Lau et al, 2019). The listening effort expended by listeners in the current study, therefore, could reflect in part the effort required to utilize binaural cues to separate target from masker speech.

It is hoped that these results add to the literature on auditory masking by helping to shed light on additional differences that EM vs IM may have on the listener, specifically what achieving a certain level of performance requires of the listener (or does to the listener). As discussed earlier, listening effort is under top-down control and can be affected by factors such as how motivated the listener is, how frustrated they are, and how much success they believe they are having. In real-world situations, the task of selectively attending to and processing target speech in the presence of masking is one that listeners are often highly motivated to solve. Listening to a friend tell an entertaining story in a busy restaurant, attending to an important intercom announcement in a crowded venue, or understanding what a family member is asking you over the phone when you are standing in the grocery store checkout line are all tasks for which we are generally willing to put forth effort. The finding that more effort may be required when masker consist of intelligible speech has implications for understanding the effects of different types of everyday listening conditions on the listener.

ACKNOWLEDGMENTS

This work was funded by the National Institutes of Health/National Institute for Deafness and Other Communication Disorders (K99DC018829 to SV). The authors acknowledge and thank Andy Byrne, Luke Baltzell, Lorraine Delhorne, Judy Dubno, Chris Mason, Jonathan Peelle, and Jan Rennies-Hochmuth for their insights on this project.

REFERENCES

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*(5), 1084. http://dx.doi.org/10.1097/AUD.0000000000000697

Arbogast, T. L., Mason, C. R., & Kidd Jr, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2169-2180. http://dx.doi.org/10.1121/1.1861598

Bönitz, H., Lunner, T., Finke, M., Fiedler, L., Lyxell, B., Riis, S. K., Ng, E., Valdes, A. L., Büchner, A., & Wendt, D. (2021). How do we allocate our resources when listening and memorizing speech in noise? A pupillometry study. *Ear and Hearing*, *42*(4), 846-859. http://dx.doi.org/10.1097/AUD.0000000000001002

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, *111*(6), 2801-2810. http://dx.doi.org/10.1121/1.1479152

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical society of America*, *25*(5), 975-979. http://dx.doi.org/10.1121/1.1907229

Conroy, C., & Kidd Jr, G. (2021). Informational masking in the modulation domain. *The Journal of the Acoustical Society of America*, *149*(5), 3665-3673. http://dx.doi.org/10.1121/10.0005038

Delorme, A., & Makeig, S. (2004). EEGLAB-Open Source Matlab Toolbox for Electrophysiological Research. *Journal of Neuroscience Methods*, *134*, 9-21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, *9*(1), 1-10. http://dx.doi.org/10.1038/s41598-019-47643-1

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both?. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), e1514. http://dx.doi.org/10.1002/wcs.1514

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behavior Research Methods*, *52*(5), 2232-2255. http://dx.doi.org/10.3758/s13428-020-01374-8

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *Journal of Neuroscience*, *30*(2), 620-628. http://dx.doi.org/10.1523/JNEUROSCI.3631-09.2010

Kidd, G., & Colburn, H. S. (2017). Informational masking in speech recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The Auditory System at the Cocktail Party*, New York: Springer, 75-109. http://dx.doi.org/10.1007/978-3-319-51662-2_4

Kidd, G. Jr., Mason, C.R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In Yost, W. A., Popper, A. N., and Fay, R. R. (Eds), *Auditory Perception of Sound Sources*, New York: Springer, 143-190. http://dx.doi.org/10.1007/978-0-387-71305-2_6

Kidd Jr, G., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., & Best, V. (2016). Determining the energetic and informational components of speech-on-speech masking. *The Journal of the Acoustical Society of America*, *140*(1), 132-144. http://dx.doi.org/10.1121/1.4954748.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291-300. http://dx.doi.org/10.1097/AUD.0b013e3182310019

Lau, M. K., Hicks, C., Kroll, T., & Zupancic, S. (2019). Effect of auditory task type on physiological and subjective measures of listening effort in individuals with normal hearing. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1549-1560. http://dx.doi.org/10.1044/2018_JSLHR-H-17-0473

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94-106. http://dx.doi.org/10.3758/s13428-017-1007-2

Mattys, S.L., Davis, M.H., Bradlow, A.R. and Scott, S.K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953-978. http://dx.doi.org/10.1080/01690965.2012.705006

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., ... & Graham, P. L. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology*, *7*, 745. http://dx.doi.org/10.3389/fpsyg.2016.00745

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., De Lissa, P., Graham, P., & Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, *21*, 1-13. http://dx.doi.org/10.1177/2331216517706396

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. *Hearing Research*, *365*, 90-99. http://dx.doi.org/10.1016/j.heares.2018.05.003

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011(1), 156869. http://dx.doi.org/10.1155/2011/156869

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*, 2642–2653. http://dx.doi.org/10.1121/1.4897398

Rennies, J., Best, V., Roverud, E., and Kidd, G. Jr. (2019). Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and listening effort, *Trends in Hearing*, *23*, 1-21 http://dx.doi.org/10.1177/2331216519854597

Rowland, S. C., Hartley, D. E., & Wiggins, I. M. (2018). Listening in naturalistic scenes: what can functional near-infrared spectroscopy and intersubject correlation analysis tell us about the underlying brain activity?. *Trends in Hearing*, *22*, 1-18. http://dx.doi.org/10.1177/2331216518804116

Stone, M. A., & Moore, B. C. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967-1977. http://dx.doi.org/10.1121/1.4868392

Versfeld, N. J., Lie, S., Kramer, S. E., & Zekveld, A. A. (2021). Informational masking with speech-on-speech intelligibility: Pupil response and time-course of learning. *The Journal of the Acoustical Society of America*, *149*(4), 2353-2366. http://dx.doi.org/10.1121/10.0003952

Visentin, C., Valzolgher, C., Pellegatti, M., Potente, P., Pavani, F., & Prodi, N. (2021). A comparison of simultaneously-obtained measures of listening effort: pupil dilation, verbal response time and self-rating. *International Journal of Audiology*, 1-13. http://dx.doi.org/10.1080/14992027.2021.1921290

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. *Hearing Research*, *369*, 67-78. http://dx.doi.org/10.1016/j.heares.2018.05.006

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 in Hearing*, *22*, 1-32. http://dx.doi.org/10.1177/2331216518800869

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends in Hearing*, *25*, 1-26. http://dx.doi.org/10.1177/23312165211027688

Zhang, M., Siegle, G. J., McNeil, M. R., Pratt, S. R., & Palmer, C. (2019). The role of reward and task demand in value-based strategic allocation of auditory comprehension effort. *Hearing Research*, *381*, 107775. http://dx.doi.org/10.1016/j.heares.2019.107775