Received: 13 August 2021 | Revised: 15 January 2022 | Accepted: 19 January 2022

DOI: 10.1111/ejn.15616

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort

Stephanie Haro[1,2] | Hrishikesh M. Rao[1] | Thomas F. Quatieri[1,2] | Christopher J. Smalt[1]

[1]Human Health and Performance Systems, MIT Lincoln Laboratory, Lexington, Massachusetts, USA

[2]Speech and Hearing Bioscience and Technology, Harvard Medical School, Boston, Massachusetts, USA

Correspondence
Christopher J. Smalt, Human Health and Performance Systems, MIT Lincoln Laboratory, Lexington, MA 02421, USA.
Email: Christopher.Smalt@ll.mit.edu

Funding information
National Institute of Health, Grant/Award Number: T32 Trainee Grant No 5T32DC000038-27; National Science Foundation, Grant/Award Number: GFRP Grant No. DGE1745303; Under Secretary of Defense for Research and Engineering, Grant/Award Number: Air Force Contract No. FA8702-15-D-0001

Edited by: Ali Mazaheri

Abstract

Everyday environments often contain distracting competing talkers and background noise, requiring listeners to focus their attention on one acoustic source and reject others. During this auditory attention task, listeners may naturally interrupt their sustained attention and switch attended sources. The effort required to perform this attention switch has not been well studied in the context of competing continuous speech. In this work, we developed two variants of endogenous attention switching and a sustained attention control. We characterized these three experimental conditions under the context of decoding auditory attention, while simultaneously evaluating listening effort and neural markers of spatial-audio cues. A least-squares, electroencephalography (EEG)-based, attention decoding algorithm was implemented across all conditions. It achieved an accuracy of 69.4% and 64.0% when computed over nonoverlapping 10 and 5-s correlation windows, respectively. Both decoders illustrated smooth transitions in the attended talker prediction through switches at approximately half of the analysis window size (e.g., the mean lag taken across the two switch conditions was 2.2 s when the 5-s correlation window was used). Expended listening effort, as measured by simultaneous EEG and pupillometry, was also a strong indicator of whether the listeners sustained attention or performed an endogenous attention switch (peak pupil diameter measure [$p = 0.034$] and minimum parietal alpha power measure [$p = 0.016$]). We additionally found evidence of talker spatial cues in the form of centrotemporal alpha power lateralization ($p = 0.0428$). These results suggest that listener effort and spatial cues may be promising features to pursue in a decoding context, in addition to speech-based features.

Abbreviations: AAD, auditory attention decoding; ANOVA, analysis of variance; EEG, electroencephalography; ERSP, event-related spectral perturbation; GLHT, general linear hypothesis test; ICA, independent component analysis (ICA); LIPSP, left inferior parietal supramarginal part; MPD, mean pupil diameter; RTPJ, right temporoparietal junction; SEM, standard error of the mean.

KEYWORDS
alpha lateralization, auditory attention decoding, competing speech, event related spectral perturbation, simultaneous talkers

1 | INTRODUCTION

Everyday listening situations often contain multiple competing talkers and listeners must engage auditory attention to focus onto one source. In voluntary sustained attention, an endogenous process, the listener directs their attention towards a source and top-down mechanisms influence how the source is represented in the cortex (Golumbic et al., 2013; Posner et al., 1984). In most environments, however, listeners do not sustain their attention to one talker continuously. Sources vie to exogenously capture listener attention, employing bottom-up processes once successful (Posner et al., 1984). Listeners may also switch their attention endogenously, shifting their attention between sources at their discretion. Listeners using auditory attention enhancement in these complex scenes would value amplification of their desired acoustic source such as a source that is being endogenously sustained or a new source that they want to endogenously switch to. Enhancement would ideally reduce the effect irrelevant stimuli have on a listener's auditory attention.

Distinct brain regions have been shown to be involved in endogenous auditory attention (Hill & Miller, 2010; Lee et al., 2013; Larson & Lee, 2014). Their studies have focused on characterizing attention between location and pitch, two core features that differ between sources in cocktail-party scenarios. The frontal-parietal region was found to be activated during endogenous auditory attention towards sources that differ in both space and pitch (Hill & Miller, 2010). Next, the frontal eye field region was found to be activated in preparation for and during endogenous attention towards sources (Lee et al., 2013). Distinct parietal activations during endogenous switches were then characterized (Larson & Lee, 2014). The right temporoparietal junction (RTPJ) and the left inferior parietal supramarginal part (LIPSP) were active during switches between sources that differed in space and pitch respectively. These protocols used small speech tokens such as alphabetic characters and unrelated sentences from a corpus (Hill & Miller, 2010; Lee et al., 2013; Larson & Lee, 2014). The regions involved with switching attention naturally between continuous speech sources have yet to be characterized and will likely recruit a combination of the previously mentioned brain regions.

Auditory attention occurs in environments that are more complex than during conventional clinical hearing assessments. Real scenes often involve multiple speech sources and reverberation that recruit speech-specific auditory processes (Liberman et al., 2016). This complexity may provide a suite of cues that can be leveraged by the listener during auditory attention. Realistic listening in contrast to clinical assessments consists of longer listening tasks that may lead to more opportunities to latch attention, greater overall comprehension due to the continuous speech context, and more listener fatigue. Identifying signals and features engaged in naturalistic switching can potentially be used to track attention states. These states can then be used to control stimuli enhancement which can improve the listener's experience. Altering the relative levels of attended and ignored stimuli can reduce listening effort and enhance attended stimuli entrainment (Mirkovic et al., 2019; Presacco et al., 2019; Seifi Ala et al., 2020). Speech enhancement has the capacity to improve listener quality of life in individuals of all ages and levels of hearing loss (Ciorba et al., 2012; Griffin et al., 2019; Liberman et al., 2016).

Auditory attention decoding (AAD) describes the process of using cortical recordings to identify to whom a listener is attending when multiple talker sources are competing for the listener's attention. AAD in combination with speaker separation has the potential to be incorporated into cognitively controlled hearing aids to provide auditory enhancement in speech-rich scenes that traditional hearing aids struggle with (Borgström et al., 2021; Popelka & Moore, 2016). The majority of these studies' protocols ask listeners to sustain attention, not invoking switches in attention (Geirnaert et al., 2021). However, it is critical to study attention switching given the prevalence of switching in real-world conditions. Various speech features and cortical recording modalities have been used to encode and decode attended stimuli (Akram et al., 2016; Ciccarelli et al., 2019; Ding & Simon, 2012; Mesgarani & Chang, 2012; O'Sullivan et al., 2015; O'Sullivan et al., 2017; Puvvada & Simon, 2017). These decoding algorithms have relied on reactive decoding of the already attended stimuli, creating a lag in the enhancement they could provide. Endogenous switches are associated with top-down attentional preparatory activity in contrast to exogenous attention switches

(Lee et al., 2013), and thus tracking the preparatory activity involved in endogenous attention switches might aid in faster enhancement in comparison to reactive decoding. Identifying preparatory features that accompany attention switches could aid in more robust auditory enhancement when combined with attended stimuli decoding.

Recent work has begun incorporating switches in their attention decoding protocols to explore alternative attention modelling techniques (Akram et al., 2016; Miran et al., 2018, 2020; Teoh & Lalor, 2019). In some auditory switching studies, the switch time was determined by the protocol (Akram et al., 2016; Hill & Miller, 2010; Lee et al., 2013; Larson & Lee, 2014; Teoh & Lalor, 2019). These studies direct attention switching using acoustic cues - directing the listener to switch sources when a gap in the stimulus occurs (Akram et al., 2016) or instructing the listener to switch attended talker location in order to track a dynamic talker (Teoh & Lalor, 2019). However when naturalistic endogenous attention switches are studied, the moment the switch occurred is known by the listener and must be extracted. There have been some attempts to obtain this switch time. For example, a button press has been used to record endogenous switch time (Miran et al., 2018, 2020). Unfortunately, a button press may create switch-locked pre-motor planning and muscle artefacts in the data, potentially confounding endogenous switching feature interpretation (Johari et al., 2019; Stephen, 2019).

In this study, we investigated two variants of endogenous switches of sustained attention between competing multitalker sources. This volitional type of attention switch is the focus of this work because listeners desire a source of their choosing to be enhanced and want to limit the effect exogenous stimuli have on their auditory attention. In this study, listeners were asked to remember when they endogenously switched using a clock in order to remove an explicit evoked response, for example, a motion artefact from a button press. For the first analysis, we performed regularized least-squares decoding of the attended talker envelope (Crosse et al., 2016). We demonstrated decoder behaviour on data that contains natural attention switches and additional realistic higher-order processes of memorization and decision making that were incorporated into the protocol. Next, we quantified the effort involved with endogenous switching using measures of EEG alpha power and pupil diameter (Seifi Ala et al., 2020). Lastly, we analysed alpha power activity related to the relative locations of the attended and unattended talker locations (Deng et al., 2020).

2 | METHODS

2.1 | Experimental protocol

Ten native English speakers (5F, 5M), with self-reported normal hearing, participated in this study. They provided informed consent to an experimental protocol that was approved by the MIT Committee on the Use of Humans as Experimental Participants and The U.S. Army Medical Research and Development Command, Human Research Protection Office. Participants were asked to sit in a sound treated booth between two loudspeakers positioned 6 feet away at ±45°. The left and right loudspeakers presented male talkers reading "Twenty Thousand Leagues Under the Sea" and "Journey to the Center of the Earth" audiobooks, respectively (O'Sullivan et al., 2019). We simultaneously recorded participant EEG and pupillometry using a dry electrode EEG system (Wearable Sensing DSI-24, Fs = 300 Hz) and eye tracking glasses (SMI ETG2, Fs = 120 Hz), respectively. A display situated in front of the participant displayed various stages of the protocol.

Figure 1 diagrams the instructional stages of a trial and the time course of the three experimental conditions. This protocol consisted of 60 one-minute trials (20 trials of each experimental condition). At the beginning of each trial, the display presented the trial task—a combination of the experimental condition and initial attended talker (left or right). Each experimental condition had an equal number of trials that began with attention to the left and right talker. The experimental condition presentation order was randomized and determined using MATLAB's uniformly distributed pseudorandom integer generator. During a trial, after approximately 30 s of attention towards the initial attended talker, listeners performed one of three tasks. Depending on the indicated experimental condition, listeners were to switch attention at their own discretion (at-will), switch attention at a directed time (directed), or not switch attention at all (sustained). To record the attention switch without the use of a button press, all three experimental conditions incorporated a time memorization task that used the visually presented elapsed time. The subject then reported this time after the end of a 1-min listening period (trial). From the onset of the trial, the display presented the elapsed time which was updated once per second. This update rate was selected instead of a finer resolution in order to prevent increased visual processing load and reduce the complexity of the time memorization task. The at-will switch involved an on-demand, listener-initiated switch; the listener used the clock to mentally note when they switched at their discretion. The directed switch had the listener switch at a time specified before

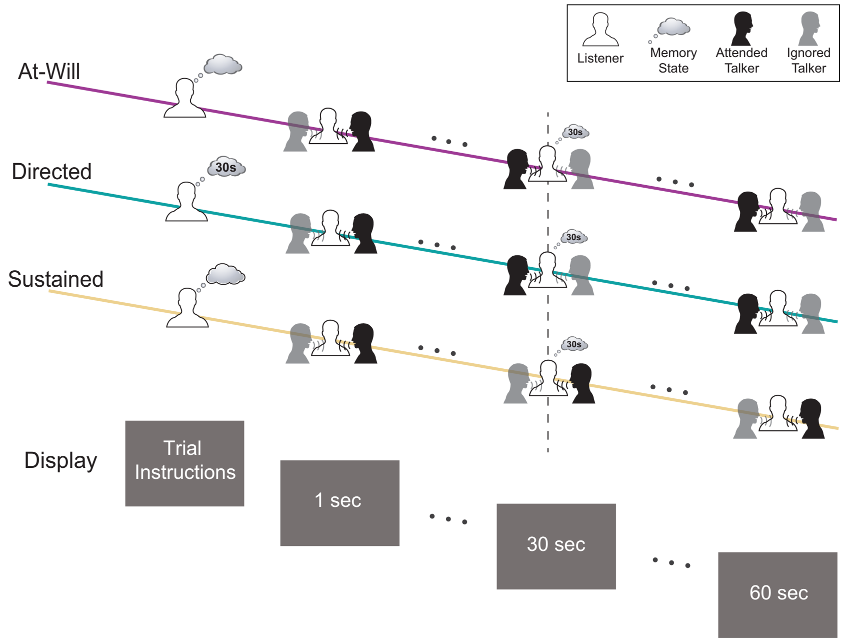

FIGURE 1 Auditory attention protocol. During the protocol, listeners were presented with two competing spatially separated audiobook stimuli. They were asked to begin each trial attending to one talker and ignoring the other. For two of the experimental conditions, listeners were asked to switch talkers approximately halfway through the trial. Listeners either switched attention at their discretion (at-will switch) or switched their attention at a time specified before the trial began (directed switch). In the third experimental condition, listeners were asked to keep their attention on the initial talker for the whole trial (sustained attention). At the end of the trial, listeners recalled the time event they either switched or continued to sustain attention

the trial began. The participant used the clock to perform the attention switch at a pre-determined time. The directed switch lacks the added online decision-making task of when to switch. It is important to note that this task is still considered endogenous since the implementation of the switch is made by the listener in contrast to having their attention exogenously captured at a given time. For the sustained condition, the listener attended to the same talker for the whole trial but was tasked to remember a time once they saw it on the clock. This task is modelled after the online decision-making that would occur in the at-will switch condition, but without the actual switch. The sustained condition controlled for the executive functioning tasks used in the at-will experimental condition (decision making and remembering time).

Figure 1 illustrates the listener memory state across the experimental conditions in the thought bubbles. All three conditions' timing events were only permitted to occur between [25,35] seconds in order to ensure ample data before and after the switch. The directed switch time was randomly generated. For the other two conditions, participants were instructed to randomize their timing events between the range of [25,35] seconds themselves. For the rest of the analysis, all trial times were normalized relative to the timing event such that it occurs at zero seconds. For simplicity, we will call this time, the switch time even though no switch occurs in the control condition where attention is sustained. Any physiological measures seen around the switch time can be attributed to executive functioning related to the listener's decision making, committing the switch time to memory, and/or switching auditory attention between spatially separated sources. Between each trial, participants recalled the trial's timing event and answered two 4-choice comprehension questions using a wireless gaming controller. Each of the 10 participant collections contain 60 min worth of EEG and pupillometry data as well as two comprehension responses for each trial. The protocol contains characteristics that should elicit measurable effort such as a reasonably difficult task that provides listener engagement and motivation (Winn et al., 2018). Listeners were asked to attend to continuous audio book speech stimuli which possibly keeps their engagement throughout the experiment more than a simpler stimuli would. Participants were motivated to follow the three experimental conditions tasks equally and value attending to the proper talker for the whole duration of the trial because they needed to answer comprehension questions from both halves of the trial.

2.2 | Auditory attention decoding

Attended and ignored talker stimuli representations differ in strength, temporal characteristics, and topography (Ding & Simon, 2012; Golumbic et al., 2013; O'Sullivan et al., 2015). These differences permit the separate talker stimuli to be distinguished from each other through the decoding of cortical signals. We performed auditory attention decoding on all three experimental conditions and used the output in two ways. First we used decoding to demonstrate that the protocol's core task of auditory attention was not severely impacted by the inclusion of visual and memory tasks. Then we used the decoding output to track a natural, nonartificial, shift in the attended talker source.

For each participant, we trained a linear decoder to predict the attended talker envelope from cortical signals (EEG). We elected to use an L2 (ridge regression) regularized least-squares decoding approach that used the attended talker envelope for training (Crosse et al., 2016). Since maximizing decoding accuracy was not the focus of the work, we chose to use a fixed regularization of 1e6 for all subjects instead of finding the listener-specific performance-maximizing regularization value via a validation process. The decoder used a sliding 500-ms window of EEG data to produce each attended talker prediction sample (Ciccarelli et al., 2019). EEG preprocessing was kept to a minimum so our decoding results could be evaluated against real-time decoding implementations that limit pre-processing for the sake of speed (Alickovic et al., 2019). The EEG data used for decoding underwent no blink rejection or visual evoked potential response pre-processing. It was then bandpassed between [2,32] Hz using EEGLab's Hamming windowed FIR filter (Ciccarelli et al., 2019; Delorme & Makeig, 2004). Each decoder is talker-invariant; training used a balanced amount of each talker's data. The ideally separated talkers' broadband audio envelopes were extracted using a nonlinear, iterative method (Horwitz-Martin et al., 2016). The bandpassed EEG and the audio envelopes were then downsampled to 100 Hz.

The training and testing process was implemented using leave-one-trial-out cross validation. The attended talker decision at a given time was determined via a Pearson correlation using a window of attended talker envelope predictions. A Pearson correlation was performed between the candidate speech envelopes and the attended speech envelope prediction, denoted by $\widehat{env}$. We evaluated the decoder with correlation windows of 10 and 5 s. In the results, we present a detailed overview of decoder performance using a length of 5 s because it provides the opportunity for comparison with the listening effort analyses that also use a 5-s window.

To evaluate the accuracy of a fold of the decoder, the decoder output, $\widehat{env}$, was first correlated with the attended and unattended talker envelopes, env_{Att} and env_{Una} (Equations (1) and (2)). In the case of the switch conditions, the attended and unattended talker envelopes are composed of concatenated envelopes from the two talkers. Attended talker decoding is considered successful when the correlation with env_{Att} is greater than env_{Una}, that is, when the decision vector, $corrDiff_{Att-Una}$, is greater than 0. Decoding accuracy is defined as the fraction of nonoverlapping time, the time-varying correlation-based decision vector, $corrDiff_{Att-Una}$ is greater than 0 (Equation 3).

$$corr_{Att} = corr(\widehat{env}, env_{Att}), \tag{1}$$

$$corr_{Una} = corr(\widehat{env}, env_{Una}), \tag{2}$$

$$corrDiff_{Att-Una} = corr_{Att} - corr_{Una}. \tag{3}$$

To characterize the shift in attention rather than the accuracy, we correlated the decoder output with the two talkers envelopes, regardless of listeners attention during the trial. Specifically, the decoder output, $\widehat{env}$, was correlated with the attended and ignored talker envelopes that the listener commenced the trial with, $env_{T\ 1}$ and $env_{T\ 2}$, respectively (Equations 4 and 5). The difference between these two correlations, denoted by $corrDiff_{T\ 1-T\ 2}$, changes sign when a switch occurs, indicating that the attended talker is no longer the talker the listener commenced the trial with (Equation 6). For the switch conditions, $corrDiff_{T\ 1-T\ 2}$ is identical to $corrDiff_{Att-Una}$ before the switch time. For the sustained condition, $corrDiff_{T\ 1-T\ 2}$ is identical to $corrDiff_{Att-Una}$ across all time.

$$corr_{T1}(t) = corr(\widehat{env}, env_{T1}), \tag{4}$$

$$corr_{T2} = corr(\widehat{env}, env_{T2}), \tag{5}$$

$$corrDiff_{T1-T2} = corr_{T1} - corr_{T2}. \tag{6}$$

2.3 | Pupillometry analysis

Pupil diameter is one measure of expended effort that we simultaneously measured during our auditory attention protocol. We performed peak-based blink detection on the raw pupil diameter data, interpolated data points containing blink artefacts, and smoothed the data using a 1-s median filter. The pupil diameter used for analysis was defined as the average pupil diameter between the left and right pupil channels. Mean pupil diameter (MPD) is a measure of pupil dilation that is normalized on a trial basis using a baseline from the onset of the trial (Equation 7). Pupil dilation may be sensitive to factors unrelated to the experimental task such as engagement, arousal, anxiety, and lighting conditions (van Rij et al., 2019). In Equation (7), D, is the pupil diameter averaged across a given 5-s window. Subscripts B and t indicate whether average pupil diameter was computed across a baseline window between [−25:20] seconds relative to the switch time or a sliding 5-second window whose latter edge spans [−20:25] seconds relative to the switch time, respectively. At 5 s for example, MPD captures the activity between [0,5] seconds proceeding the switch, not just the activity at 5 s.

$$MPD(t) = \frac{D_t - D_B}{D_B} * 100. \tag{7}$$

We hypothesized that peak MPD around the switch would be modulated by the amount of effort required by the condition's set of tasks. MPD peak magnitude around the switch was automatically detected and used for statistical testing. To test the effect of experimental condition on peak MPD, we ran a two-factor analysis of variance (ANOVA) with experimental condition modelled as a factor and participant modelled as a random factor. In-addition, we performed planned, pairwise *t*-tests between experimental conditions with a Bonferroni correction that conservatively adjusts the *t*-test *p* values to correct for type-1 error that may arise during the simultaneously run *t*-tests (Armstrong, 2014).

2.4 | EEG analysis

2.4.1 | Event-related spectral perturbation

In previous work, cortical measures that were already being recorded for decoding have been used to quantify listening effort during sustained attention (Seifi Ala et al., 2020). To investigate effort during an endogenous attention switch, we also evaluated EEG measures of effort around the switch. The EEG data was preprocessed differently for this banded analysis than was done for the least-squares decoding analysis. In contrast to single trial decoding, EEG power band analysis is sensitive to blinks. Blink artefacts were removed from the data using independent component analysis (ICA) methods found in the EEGlab toolbox (Delorme & Makeig, 2004). Instead of absolute alpha band power, we computed a measure of relative alpha power in the form of event-related spectral perturbation (ERSP) (Makeig, 1993). Similar to MPD, ERSP is a measure of alpha normalized on a trial basis using a baseline window at the beginning of the trial. *ERSP*(t, c) is a function of both time, t and EEG channel, c, (Equation 8). The absolute alpha power, P, was computed as the sum of squared spectral density values between [8,12] Hz. Spectral density was computed across each analysis window using MATLAB's pwelch method. The baseline window, B, indicates that the spectral power was computed across a baseline window between [−25:20] seconds relative to the trial's switch time (Equation 8). The variable, t, indicates that the spectral power was computed on a sliding 5-s window whose latter edge spans [−20:25] seconds relative to the switch time. Both MPD and alpha ERSP were computed every 10 ms with a 4.99-s analysis window overlap. ERSP was computed individually for each channel, c, using that channel's baseline alpha power.

$$ERSP(t,c) = \frac{P(t,c) - P(B,c)}{P(B,c)} * 100. \tag{8}$$

In an attempt to remove muscle artefacts that were not resolved by ICA, ERSP samples that were 1.5 times the interquartile range beyond the third and first quartile were removed and replaced with linearly interpolated values (Elliott & Woodward, 2007). We computed the mean ERSP across the parietal subset of channels to arrive at the parietal ERSP measure. In addition to the baseline window normalization, we *z*-scored trial-level MPD and ERSP within each participant to highlight experimental condition differences instead of participant differences. We applied a smoothing low-pass filter on alpha ERSP order to aid in the automatic detection of the ERSP minimum near the switch time. We hypothesized that minimum parietal alpha ERSP around the switch would also be modulated by the relative effort required by the different condition's set of tasks. To test the effect of experimental condition on minimum parietal alpha ERSP, we ran a two-factor ANOVA with experimental condition modelled as a factor and participant modelled as a random factor. To test differences in parietal alpha ERSP between conditions, we performed a planned, pairwise *t*-tests between experimental conditions with a Bonferroni correction.

In addition to investigating parietal alpha ERSP, we performed exploratory testing across three power bands (delta, theta, alpha) and three channel subsets (frontal, centrotemporal, and parieto-occipital). These additional ERSP responses underwent the same artefact rejection, *z*-scoring, smoothing, and automatic minimum detection steps as parietal alpha ERSP. Nine separate two-factor ANOVA tests were run on the nine minimum ERSP measures with experimental condition modelled as a factor and participant modelled as a random factor. These ANOVA *p* values were corrected for multiple comparisons using the false-discovery-rate (FDR) method (Benjamini & Hochberg, 1995). Two separate FDR corrections were considered that differ in how conservative the correction is, false positive rates (FPRs) of 0.1 and 0.05 were applied. We also evaluated two null hypotheses that tested whether the sustained ERSP measure was equal to each switch condition's ERSP measure. The separate null hypotheses were FDR corrected separately using FPRs of 0.1 and 0.05.

2.4.2 | Lateralized alpha event related spectral perturbation

In addition to listening effort, attended and suppressed talker spatial cues may also modulate alpha power during auditory attention (Deng et al., 2020). To assess this phenomenon in our data, we extracted an alpha feature (ERSP magnitude from Equation 8) that highlights hemispheric differences in response to the relative talker locations. In contrast to the decoding and listening effort analyses performed on a sliding window, we elected to perform the alpha lateralization analysis across long segments of time (i.e., at the trial level). Each trial was partitioned into two 20-s segments, located before and after the switch time. Each condition has forty 20-s segments; twenty segments correspond to left talker attention and twenty segments correspond to right talker attention. The choice to segment the trials was inspired by previous work that measured alpha lateralization over the entirety of their trials (Deng et al., 2020). Additionally, using a segment's worth of samples permits a better measure of alpha modulation due to spatial cues since alpha is also modulated by effort in this protocol.

The alpha lateralization analysis steps in Equations (9)–(13) extract the spatial cues in the topography, and quantify the alpha lateralization in particular brain regions. This process was implemented separately for each experimental condition. First, alpha ERSP was averaged across the segments, m, that correspond to time when the left talker was attended, denoted by $A_L(t, c)$ where t is time and c is channel (Equation 9). Alpha ERSP was averaged separately across segments that correspond to right talker attention, denoted by $A_R(t, c)$ (Equation 10). To highlight alpha's capacity to reflect relative talker spatial cues, the difference was taken between the left and right attended talker alpha ERSP responses, $A_L(t, c)$ and $A_R(t, c)$, and averaged across the 20-s long segment duration, denoted by $A_{L-R}(c)$ (Equation 11). When $A_{L-R}(c)$ is visualized as a topography, channels with positive magnitudes indicate a larger alpha synchronization to a left-located attended talker than a right-located unattended talker.

$$A_L(t,c) = \frac{1}{20}\sum_{m=1}^{20} ERSP(t,c), \quad m \in \text{segments with left talker attention}, \tag{9}$$

$$A_R(t,c) = \frac{1}{20}\sum_{m=1}^{20} ERSP(t,c), \quad \text{m} \in \text{segments with right talker attention}, \tag{10}$$

$$A_{L-R}(c) = \frac{1}{20}\int_{t=0}^{20} A_L(t,c) - A_R(t,c)dt. \tag{11}$$

To quantify a given region's capacity to reflect alpha lateralization, $A_{L-R}(c)$ was averaged across a given subset of channels separately for each hemisphere (Equations 12 and 13). Because these lateralization measures rely on the net left minus right attended talker alpha response, the ipsilateral and contralateral distinctions are relative to the left talker being attended. Therefore, a stronger A_{ipsi} indicates a stronger alpha ERSP response in the hemisphere ipsilateral to the left attended talker (Equation 12). A stronger A_{contra} indicates a stronger alpha ERSP response in the hemisphere contralateral to the left attended talker (Equation 13). Previous work has shown larger alpha magnitude lateralization in the hemisphere ipsilateral to the attended source (contralateral to the ignored source). This alpha lateralization may reflect suppression of the ignored source (Deng et al., 2020).

$$A_{ipsi} = \frac{1}{C}\sum_{c=1}^{C} A_{L-R}(c), \text{c} \in \text{left channel subset}, \tag{12}$$

$$A_{contra} = \frac{1}{C}\sum_{C=1}^{C} A_{L-R}(c), \text{c} \in \text{right channel subset}. \tag{13}$$

As indicated by previous work, we hypothesized that alpha lateralization in the parieto-occipital region would be sensitive to the relative spatial locations of the attended and unattended talkers (Deng et al., 2020). A three-factor ANOVA was run on hemispheric alpha magnitudes with experimental condition and the spatial location (i.e., ipsilateral vs. contralateral) modelled as factors and participant modelled as a random factor. We hypothesized that spatial location would have a differential effect on each hemisphere's alpha magnitude, indicating alpha lateralization.

In addition to parieto-occipital region, we explored alpha lateralization in the centrotemporal region and across the entire hemisphere. We performed the same three-factor ANOVA test on each of the additional regional alpha lateralization measures. To correct for the exploratory ANOVA tests performed on these additional regions, we applied a Bonferroni correction since the number of exploratory tests did not warrant an FDR correction.

3 | RESULTS

3.1 | Attended talker comprehension

Participants answered 120 4-choice comprehension questions and scored above chance (25%) with a mean accuracy of 56% (SEM = 3%). Our protocol's lack of attended talker continuity may be the main reason our comprehension accuracy is lower than another study that used the same stimuli and a larger fraction of the comprehension question corpus (O'Sullivan et al., 2019). Our lower accuracy most likely can be attributed to the fact that our attended talker was randomly assigned at the onset of every trial and two-thirds of the time, the attended talker was switched midway through the trial. The lower comprehension scores may also be due to the additional visual, decision-making, and working-memory tasks that listeners are asked to perform while they attend to auditory stimuli (O'Sullivan et al., 2019; Senkowski et al., 2008). Participants achieved mean comprehension accuracies of 58%, 51%, and 58% across at-will, directed, and sustained conditions, respectively (Figure 2a). A two-factor ANOVA was run with experimental condition modelled as a factor and participant modelled as a random factor. It determined a main effect of experimental condition on comprehension accuracy ($F_{(2,18)} = 4.247, p = 0.0308$). However, Bonferroni-corrected pairwise *t*-tests found no differences in comprehension accuracy between conditions.

3.2 | Attended talker decoding

Attention decoding was evaluated using nonoverlapping correlation window lengths of 10 and 5 s (Equation 3). The grand-mean accuracy dropped from 69.4% (SEM = 1.9%) to 64.0% (SEM = 1.5%) when the correlation window length was halved from 10 to 5 s. Both the 10 and 5-s correlation window data illustrate similar relative accuracies across experimental conditions. For simplicity, we will present the 5-s correlation window data since it shares the same duration as the window used for the MPD and ERSP analyses. A two-factor ANOVA was run with experimental condition modelled as a factor and participant modelled as a random factor. It found no effect of experimental condition on trial-level decoding accuracy evaluated using the 5-s correlation window ($F_{(2,18)} = 1.83, p = 0.189$) (Figure 2b).

In addition to computing the attended talker decoding accuracy, we computed a similarity metric between the decoder output and the initial attended and unattended talkers (Equation 6). When a listener engaged in a switch in attention, it took approximately half the length of the correlation window for the correlation with the initial talker to weaken below the correlation with the secondary talker (Figure 3). This result was present for both the 10- and 5-s correlation window data. Again, for simplicity we are only reporting the 5-s data results. For the 5-s correlation window, grand-mean $corrDiff_{T\,1\,-\,T\,2}$, changed sign at 2.31 and 2.15 s for the at-will and directed switch conditions, respectively.

3.3 | Event related spectral perturbation and mean pupil diameter

The grand-mean alpha ERSP topographic distribution for each experimental condition is visualized in Figure 4. Around the switch time, grand-mean alpha ERSP topographies demonstrate differences in the conditions that contain a switch in contrast to the sustained condition. The grand-mean MPD and parietal alpha ERSP time course with standard error of the mean (SEM) is depicted in Figure 5a,b. All conditions across the two modalities have a slow trend in magnitude over the course of the trial, similar to Seifi Ala et al. (2020). Before the switch

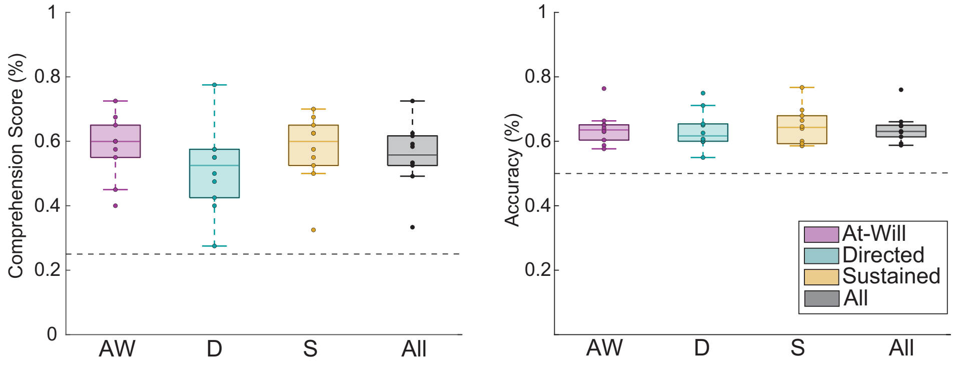

FIGURE 2 Comprehension and decoding accuracy across experimental conditions. (a) Participants achieved a mean accuracy of 56% (SEM = 3%) on forced-choice comprehension questions (chance = 25%) asked at the end of each trial. There was a main effect of experimental condition on comprehension but no pairwise differences. (b) Least-squares attended talker decoding achieved an accuracy of 64.0% (SEM = 1.5%) when computed over a nonoverlapping 5-s correlation window. There are no differences in accuracy between experimental conditions

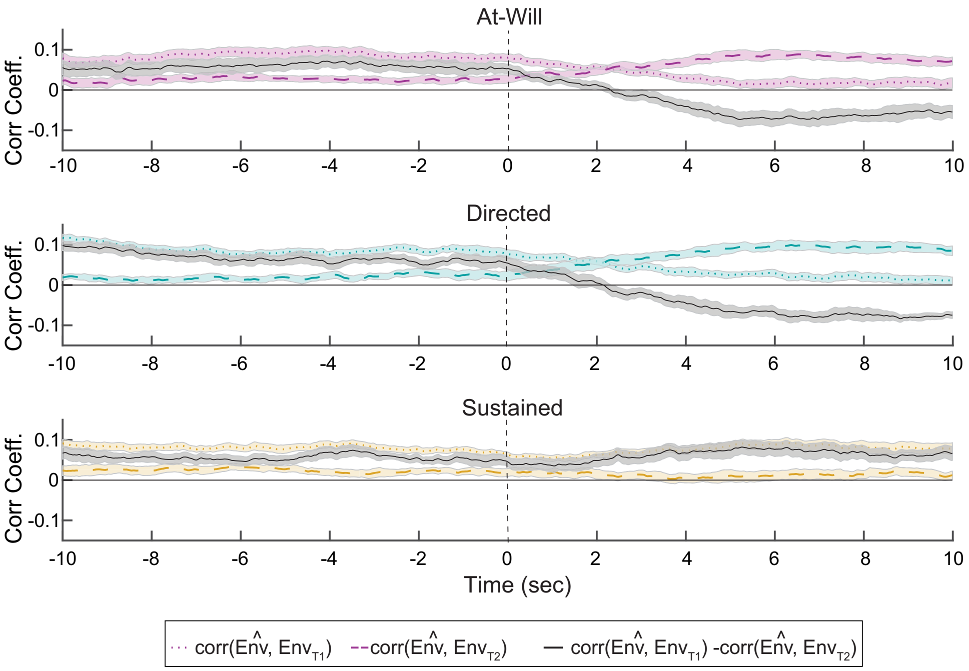

FIGURE 3 Least-squares decoding illustrates smooth switches in attention. In order to visualize the switch in talkers, the decoder output was correlated with the trial's initial attended and unattended talker speech envelopes. In the experimental conditions that contain a switch between talkers (top two panels), the correlation with the respective talker envelopes flip sign at a lag of approximately half the correlation window

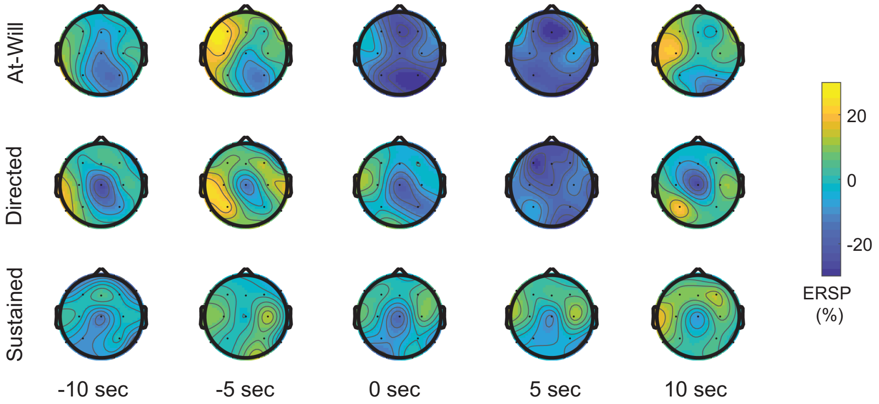

FIGURE 4 Grand-mean alpha event related spectral perturbation (ERSP) topography for each experimental condition sampled at time points before, during, and after the switch time. Alpha ERSP was computed using a sliding 5-s window of data, therefore the sampled topographies shown capture activity from the preceding 5 s of time. At 5 s, the switch conditions (top two panels) have larger alpha desynchronizations than the sustained condition

time, both MPD and parietal alpha ERSP appear similar across experimental conditions, likely because the attention task is similar across the trials in that time region. Near the switch time, both MPD and parietal ERSP fluctuate and separate by experimental condition. The peak MPD and minimum parietal alpha ERSP across the three conditions are visualized in Figure 5a,c.

A two-factor ANOVA was run separately for the peak MPD and minimum parietal alpha ERSP measures with experimental condition modelled as a factor and participant modelled as a random factor. The ANOVA found a main effect of experimental condition on peak MPD around the switch ($F_{(2,18)} = 7.668, p = 0.0039$). The Bonferroni-corrected pairwise *t*-tests found a difference in the peak MPD measure between the sustained and at-will switch conditions ($p = 0.034$) (Figure 5b). A separate ANOVA found a main effect of experimental condition on minimum parietal alpha ERSP around the switch ($F_{(2,18)} = 5.715, p = 0.012$). The Bonferroni-corrected pairwise *t*-tests found a difference in the minimum parietal alpha ERSP measure between the sustained and at-will switch conditions ($p = 0.016$), (Figure 5d). No other MPD and ERSP differences between conditions were found.

Exploratory ERSP analysis was performed across nine spectral-spatial ERSP combinations made up of three power bands (delta, theta, alpha) and three channel subsets (frontal, centrotemporal, and parieto-occipital). Two separate FDR corrections were run on main-effect *p* values, that differed in their level of conservativeness (FPRs of 0.1 and 0.05 were considered). The less conservative FDR correction (FPR of 0.1) determined a main effect of experimental condition on minimum centrotemporal alpha around the switch ($F_{(2,18)} = 5.236, p = 0.0161$). The more conservative FDR with a FPR of 0.05 found a main effect of experimental

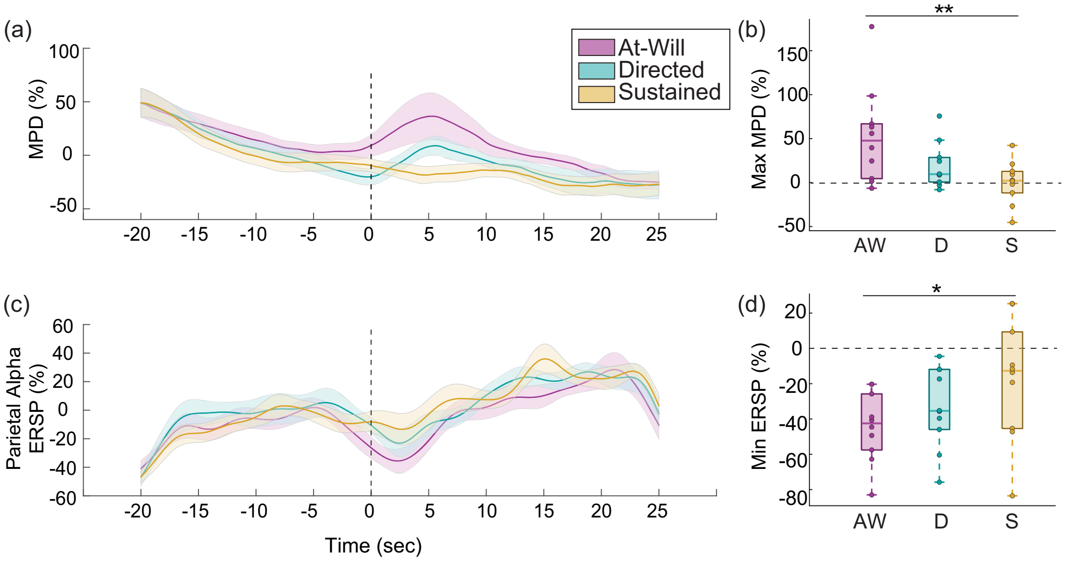

FIGURE 5 Mean pupil diameter (MPD) and parietal alpha event related spectral perturbation (ERSP). (a) Grand-mean MPD over time. (b) Peak MPD measure around the switch time. (c) Grand-mean parietal alpha ERSP over time. (d) Minimum parietal alpha ERSP measure around the switch time. Experimental condition has a main effect on peak MPD and minimum parietal ERSP around the switch time. Peak MPD and minimum parietal alpha ERSP measures are different between the sustained attention and at-will switch conditions

condition on parieto-occipital alpha ERSP ($F_{(2,18)} = 7.225, p = 0.00497$). To correct for multiple comparisons of pair-wise *t*-test *p* values, two separate FDR corrections were run using FPRs of 0.1 and 0.05. The more conservative FDR (FPR of 0.05) determined a difference in minimum centrotemporal alpha ERSP around the switch between the sustained and at-will switch conditions ($p = 0.0056$). Minimum parieto-occipital alpha ERSP around the switch was found to be different between the sustained and at-will switch conditions ($p = 0.0056$, FPR of 0.05). Minimum frontal alpha ERSP around the switch was found to be different between the sustained and directed switch conditions ($p = 0.0041$, FPR of 0.05). Together these results show evidence of global alpha desynchronizations around the switch time that may be due to the expended effort required to switch attended sources and/or perform a taxing decision making task while attending.

3.4 | Alpha lateralization

Figure 6a illustrates the net alpha ERSP, $A_{L-R}(n)$, for each experimental condition. We hypothesized that suppression towards the unattended talker would manifest as a stronger parieto-occipital alpha in the hemisphere ipsilateral to the net attended talker location (left hemisphere). A three-factor ANOVA revealed no effect of spatial location on parieto-occipital alpha hemispheric magnitude ($F_{(1,9)} = 3.44, p = 0.0966$). Since alpha lateralization was not present in the parieto-occipital region, we did not perform a generalized linear hypothesis test (GLHT) on the region's alpha measure.

In a secondary exploratory analysis, centrotemporal and hemispheric alpha lateralization was evaluated. Bonferroni-corrected testing found a main effect of the spatial location on centrotemporal alpha hemispheric magnitude ($F_{(1,9)} = 7.728, p = 0.0428$). Because the centrotemporal region had the strongest main effect of spatial location, centrotemporal alpha hemispheric magnitudes, A_{ipsi} and A_{contra}, are visualized for each experimental condition in Figure 6b. We performed a GLHT on centrotemporal alpha to test for within condition differences between ipsilateral and contralateral hemispheric alpha magnitudes. Although spatial location had a main effect on centrotemporal alpha lateralization, none of the GLHT comparisons were significant after a Bonferroni correction. The uncorrected GLHT found that within the experimental conditions, ipsilateral and contralateral centrotemporal alpha magnitude were different for the at-will condition ($p = 0.03991$) but not the other two conditions ($p = 0.18243$ (directed), $p = 0.1760$ (sustained)).

4 | DISCUSSION

We studied endogenous attention switching in the context of developing decoding algorithms that can be used in natural, every-day multitalker listening environments. Our experimental protocol allowed listeners to

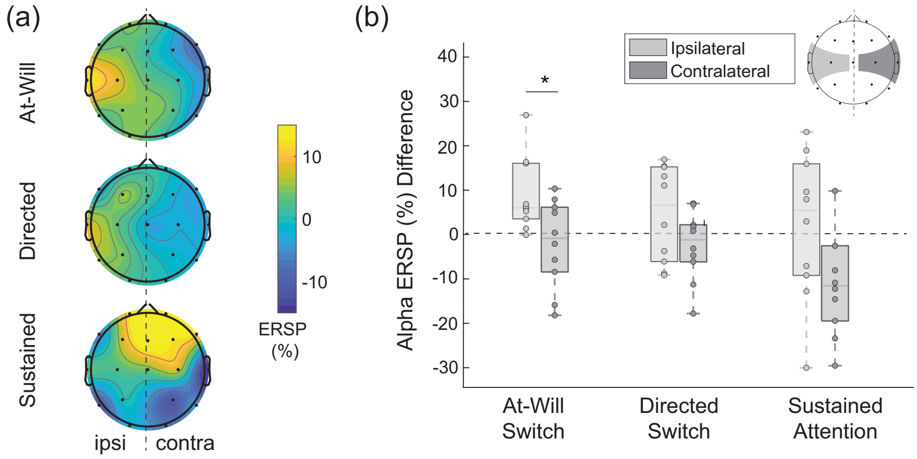

FIGURE 6 Alpha lateralization across experimental attention-switch conditions (a) Net alpha ERSP topography in response to left and right talker attention. Channels with positive magnitudes indicate a larger alpha synchronization to a left-located attended talker than a right-located unattended talker. (b) The relative centrotemporal alpha ERSP ipsilateral to, and contralateral to the attended talker derived from the topography in panel a, where a main effect was observed between hemispheres

endogenously switch attention between continuous speech sources while their effort was characterized through EEG and pupillometry measurements. In addition to effort, we detected two types of endogenous attention switches using both talkers' speech envelopes and spatial locations. While this decoding result is not the first to demonstrate endogenous attention switch decoding (Miran et al., 2018, 2020), to the best of our knowledge, it is the first study to decode multitalker continuous speech without the potential confound of sensorimotor planning. We also introduced a novel characterization of the effort involved with attention switching between speech sources. This builds upon effort measures associated with sustained attention between competing speech sources (Seifi Ala et al., 2020) and attention switching between pair of competing speech tokens (McCloy et al., 2017). Pupil diameter and parietal alpha measures indicated that the effort associated with performing an endogenous at-will attention switch was greater than our sustained-attention condition. Listener centrotemporal alpha power was also found to be modulated by the relative spatial locations of the stimuli. Our decoding results highlight latencies inherent in speech-feature decoding. Our EEG and pupil diameter findings support leveraging attention-switch decoding and other nonspeech features for improving the accuracy and decreasing the decision latency involved with cognitively controlled hearing aids.

4.1 | Switching latency of envelope-based attention decoding

Listeners who struggle with speech understanding in multitalker scenes would greatly benefit from enhancement that instantaneously cues on the talker they wish to attend. For practical applications, decoding algorithms must operate in a causal manner, incrementally producing a decoding decision from a given window of previous data. When there is a switch in attention, this analysis window length translates into a decoding latency. Although the 5-s correlation window would produce a faster detection of a switch than a 10-s correlation window, it is at a cost. The 5-s correlation decision produced noisier predictions over time and reduced decoding accuracy by 5.2% when compared to the 10-s window. Another group systematically studied this trade-off using a linear model on another data set that contained simulated attention switches. They achieved optimal accuracies of 62% and 68% using an evaluation window of 2.54 and 11.28 s, respectively (Geirnaert et al., 2019).

Our study in contrast, evaluates performance on EEG data that contains real human switches in attention and confirms that switches can be detected after a lag equal to approximately half the decision window length using a standard least-squares decoding method. We originally hypothesized that when an attention switch occurs, there is a measurable latency associated with the time it takes for the listener to go from attending to one source to another. While the decoding lag defines the fastest the decoder can detect an attended talker change, it assumes a negligible human switching delay. In both 10 and 5 s evaluations of our decoder, we observed the decision vector, $corrDiff_{T1-T2}$, change sign at a lag equal to half the respective correlation window length, indicating a switch in the listener's attended talker. For the 5-s correlation window length, mean switch time was 2.31 and 2.15, respectively, for our at-will and directed experimental conditions (Figure 3). Since the decoded switch time was less than half the window size, this indicates that listeners were potentially switching slightly before the reported switch time. Future work needs to be done to more precisely define and measure when and how long it takes for a listener to switch attention.

Supplementing envelope-decoding with other features may further reduce the algorithmic switching time for attention decoding. For context, another group's state-modeling algorithm found algorithmic delays of 1.9, 1.75, and 1.5 s for simulated switch data, real switches in EEG, and real switches in MEG, respectively (Miran et al., 2018). The decoding lag in our data set and others, demonstrates the need for further inquiry into alternative decoding features such as expended effort, cortical power lateralization in response to spatial cues, pupillometry, and eye-gaze. We propose that this multimodal decoding approach could be implemented in various ways. The data would first need to utilize a normalization or remapping step to place the various feature types into the same space (Geirnaert et al., 2020). One option is to train a model that treats the time varying pupil and EEG measures of effort as additional physiological channels that can be concatenated to the EEG data. In this case, nonlinear models such as a recurrent neural network, could leverage its capacity to retain a running memory to merge various physiological measures that operate on different time scales to produce an attended prediction (Geravanchizadeh & Roushan, 2021). A state-modeling approach could also be used to update the attended talker state when an marker of effortful listening has taken place (Miran et al., 2018). A third option is to combine a weighted sum of separate classifiers to produce a prediction that involves the various cues a listener might be leveraging. For example, decoded attended talker location and acoustic predictions could be combined and scaled with an effort-measure weight.

4.2 | Increased listening effort is associated with auditory attention switching

Our results suggest that the effort required to switch attention was greater than the effort required to sustain attention. We found a main effect of experimental condition on peak MPD and minimum parietal alpha ERSP around the switch time. Peak MPD and minimum parietal alpha ERSP reflect both the effort due to switching and higher-order cognitive tasks (in-the-moment decision making and time memorization), depending on the experimental condition instructions (Figure 5b,d). In our experiment, both the at-will and sustained conditions involve decision making and time memorization and only differ in whether an attention switch occurs. Therefore the differences in the switch and sustained condition measures are due to the effort required to implement the switch in attention. We did not find differences in peak MPD or minimum parietal alpha ERSP between the directed condition and other two conditions. This indicates that these two measures of effort are not sensitive to differences between the directed condition's tasks and the other two condition's tasks. In addition to fluctuations around the switch time, Figure 5a illustrates a slow downward trend in MPD and slow rising trend in parietal alpha ERSP. Its not clear whether these changes in MPD and ERSP are related to a change in effort and may be indicative of physiological adaptation over the course of the trial.

Our results show that pupil diameter increases during our complex attention switching tasks in manner that is consistent with previous pupil diameter measures performed during an exogenous attention switch between competing alphabetic character pairs (McCloy et al., 2017). In addition to understanding the effort associated with attention switching, pupil diameter measured throughout the entire 60-min collection can be leveraged to determine the impact a listener's effort has on decoding accuracy. In future studies, pupil diameter can be used as a measure of fatigue over the course of long stretches of effortful listening and to determine auditory training's efficacy in reducing such fatigue (Pichora-Fuller et al., 2016). These attention switching conditions could be implemented in clinic to gauge listener effort when performing auditory attention between stimuli with low speech intelligibility (Pichora-Fuller et al., 2016; Paul et al., 2021; Winn et al., 2018; Zekveld et al., 2018). These measures could help gain insight on an individual's fatigue associated with everyday difficult listening conditions out side the clinic as well.

Our alpha ERSP results are consistent with previous sustained attention effort characterization (Seifi Ala et al., 2020). We also found stronger alpha desynchronization in the at-will experimental condition which had the most demanding combination of tasks. In the at-will condition, the listener expended effort around the switch when they performed an online decision-making task of when to switch, wrote the switch time to memory, and shifted auditory attention between sources. As expended effort increases, cortical networks activate, resulting in decreased cortical synchrony and decreased alpha ERSP (Jensen & Mazaheri, 2010; Pfurtscheller, 2001; Seifi Ala et al., 2020). Our exploratory analysis found a main effect of experimental condition on minimum alpha ERSP computed across centrotemporal and parieto-occipital channels. Our results differ in the brain regions activated. These differences may be due task differences, other previous work has also shown frontal and centrotemporal region activations during endogenous attention switches (Hill & Miller, 2010; Larson & Lee, 2014).

4.3 | EEG alpha power is lateralized by attentional spatial cues

Several prior studies have suggested that the spatial location of acoustic stimuli lateralizes alpha power during an attention task (Bonnefond & Jensen, 2012; Bednar & Lalor, 2018; Deng et al., 2020; Weisz et al., 2011). It has been shown that stimulus suppression increases parieto-occipital alpha in the hemisphere ipsilateral to the attended talker in an attention task of competing sources (Deng et al., 2020). We hypothesized parieto-occipital alpha lateralization would be present did not find a main effect of spatial location on parieto-occipital alpha hemispheric magnitude ($p = 0.0966$). However, our Bonferroni-corrected exploratory analysis performed on two additional channel subsets found evidence of centrotemporal alpha lateralization. Of the three channel subsets that were evaluated, spatial location impacted centrotemporal alpha hemispheric magnitude the most ($p = 0.0428$). The channel subset where the alpha lateralization was found to be strongest may differ from previous work because our task utilized read speech rather than previous work with single syllables (Deng et al., 2020). Another possibility could be related to the fact that we computed alpha ERSP instead of individualized peak alpha magnitude for our alpha measure, or the fewer number of EEG electrodes in our study.

Our results further support that even with a demanding task of attention between continuous speech stimuli, some alpha lateralization effects may be present. Although there is evidence of spatially modulated alpha power, this cue is limited for single-trial decoding use due to the 20-s long segments over which the feature was computed. Better features may exist for leveraging spatial cues for decoding. For example, a decoding method that used common spatial pattern filters to determine directional focus without the use of speech-features, performed at an accuracy of 80% and window length of 1 s (Geirnaert et al., 2020).

4.4 | Leveraging attention switches for a cognitively controlled hearing aid

Cognitively controlled hearing aids have the capacity to improve the listener experience in cluttered environments through listener-steered speech enhancement (Geirnaert et al., 2021). Understanding endogenous switching may speed attention decoding by identifying the intended attended talker throughout a switch before the new attended talker is fully attended to. Speech-feature based decoding relies on the attended speech being encoded in the listener's cortical signals. It remains unknown how these speech-feature based algorithms would work on real attention switches in individuals with hearing impairment (Decruy et al., 2020; Van Canneyt et al., 2021). On the other hand, sensing effort expended in an attempt to attend to a new source and ignore another, could be leveraged to help decode intent in this situation. It is probable that the attention processes involved with an endogenous switch may begin to show themselves in cortical signals earlier than an exogenous capture in attention due to the decision-making and planning involved. Therefore, supplementing speech-feature based decoding with features that are directly related to switches in auditory attention, may result in decreased decoding lag and increased accuracy. The neural and pupil diameter markers associated with switching effort, as shown in our results, could potentially be leveraged as one of these features.

This work further supports exploring nonacoustic, multimodal features for attention decoding. Our results demonstrated that speech-feature based decoding still functions in the presence of additional higher-order cortical tasks, indicating that nonspeech features have promise to be fused with speech-features for robust multicue feature decoding. This work did not focus on maximizing decoding accuracy nor minimizing the switch detection lag but future work could aim to use these additional features as part of decoding models. Specifically, alpha ERSP and pupil diameter features may be relevant since their slope began to change sign slightly before or at the time listeners reported their switch. Individuals naturally also use both auditory and visual attention in a multitalker listening task, therefore eye gaze can also be pursued as a noncovert feature for auditory attention decoding (Best et al., 2017; Favre-Felix et al., 2018; O'Sullivan et al., 2019).

5 | CONCLUSION

In this study, we characterized the effort associated with endogenous auditory attention switching using both cortical and pupil diameter measures. Decoding real endogenous switches in attention illustrated the problematic lag associated with decoding methods that rely on attended talker speech features. MPD and alpha ERSP measures of effort were sensitive to endogenous switches of auditory attention. Our listening effort features have a potential application in a multimodal, multifeature decoding algorithm for use in a cognitively controlled hearing aid. Both effort features hold promise in being quick to reflect the onset of switching while being stable in their time course, potentially leading to a shorter lag in switch detection. The study's effortful attention switching tasks may also apply to the development of objective neural markers of listening effort that are intended for clinical use (Paul et al., 2021; Pichora-Fuller et al., 2016; Zekveld et al., 2018). One last application of these switching effort measures is in the field of attention disorders and development (Hanania & Smith, 2010). Characterizing auditory attention across populations and within individuals is important to pursue in combination with developing effort-based features for decoding. In addition to clinical hearing ability (Decruy et al., 2020; Fuglsang et al., 2020; Vanthornhout et al., 2018), expended cognitive effort during listening may greatly impact an individuals auditory attention decoding accuracy. Cognitive-controlled hearing-aid technology can leverage listener effort in many ways. Decoding algorithm speed and accuracy, listener benefit due to enhancement, and efficacy of auditory training can all utilize measures of effort.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. SH was supported in part by an National Institute of Health (NIH) T32 Trainee Grant No. 5T32DC000038-27 and the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. DGE1745303. The authors would like to acknowledge the organizers, the 2019 NSF-funded Neuromorphic Cognition Engineering Workshop in Telluride, CO, USA, where this protocol was conceived and pilot data were collected. We thank Edmund Lalor for providing the stimulus and comprehension questions and Pablo Ripolles for helping collect the pilot data. We would also like to thank workshop participants who participated in discussions regarding this protocol, including Malcolm Slaney, Gregory Ciccarelli, Edmund Lalor, Alain de Cheveigne, Behtash Babadi, Nima Mesgarani, and Tom Francart.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

SH: Conceptualization, Methodology, Formal Analysis, Software, Visualization, Writing-Original Draft Preparation, Writing-Review and Editing. HMR: Pupillometry Analysis, Writing-Review and Editing. TFQ: Conceptualization, Writing-Review and Editing. CJS: Conceptualization, Funding Acquisition, Methodology, Project Administration, Writing-Review and Editing.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15616.

DATA AVAILABILITY STATEMENT

The raw data, code, and stimuli can be obtained upon request to the corresponding author. Data are approved for public release. Distribution is unlimited.

ORCID

Stephanie Haro https://orcid.org/0000-0002-4972-2632
Hrishikesh M. Rao https://orcid.org/0000-0003-2754-2419
Thomas F. Quatieri https://orcid.org/0000-0003-1925-6340
Christopher J. Smalt https://orcid.org/0000-0002-3467-5888

REFERENCES

Akram, S., Presacco, A., Simon, J. Z., Shamma, S. A., & Babadi, B. (2016). Robust decoding of selective auditory attention from meg in a competing-speaker environment via state-space modeling. *NeuroImage*, *124*, 906–917.

Alickovic, E., Lunner, T., Gustafsson, F., & Ljung, L. (2019). A tutorial on auditory attention identification methods. *Frontiers in Neuroscience*, *13*, 153.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508.

Bednar, A., & Lalor, E. C. (2018). Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *NeuroImage*, *181*, 683–691.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Best, V., Roverud, E., Streeter, T., Mason, C. R., & Kidd Jr, G. (2017). The benefit of a visually guided beamformer in a dynamic speech task. *Trends in Hearing*, *21*, 2331216517722304.

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Current Biology*, *22*(20), 1969–1974.

Borgström, B. J., Brandstein, M. S., Ciccarelli, G. A., Quatieri, T. F., & Smalt, C. J. (2021). Speaker separation in realistic noise environments with applications to a cognitively-controlled hearing aid. *Neural Networks*, *140*, 136–147.

Ciccarelli, G., Nolan, M., Perricone, J., Calamia, P. T., Haro, S., OâĂŹSullivan, J., Mesgarani, N., Quatieri, T. F., & Smalt, C. J. (2019). Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods. *Scientific Reports*, *9*(1), 1–10.

Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical Interventions in Aging*, *7*, 159.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (MTRF) toolbox: A matlab toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, *10*, 604.

Decruy, L., Vanthornhout, J., & Francart, T. (2020). Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research*, *393*, 107961.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Deng, Y., Choi, I., & Shinn-Cunningham, B. (2020). Topographic specificity of alpha power during auditory spatial attention. *Neuroimage*, *207*, 116360.

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89.

Elliott, A. C., & Woodward, W. A. (2007). *Statistical analysis quick reference guidebook: With SPSS examples*: Sage.

Favre-Felix, A., Graversen, C., Hietkamp, R. K., Dau, T., & Lunner, T. (2018). Improving speech intelligibility by hearing aid eye-gaze steering: Conditions with head fixated in a multitalker environment. *Trends in Hearing*, *22*, 2331216518814388.

Fuglsang, S. A., Märcher-Rørsted, J., Dau, T., & Hjortkjær, J. (2020). Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. *Journal of Neuroscience*, *40*(12), 2562–2572.

Geirnaert, S., Francart, T., & Bertrand, A. (2019). An interpretable performance metric for auditory attention decoding algorithms in a context of neuro-steered gain control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(1), 307–317.

Geirnaert, S., Francart, T., & Bertrand, A. (2020). Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns. *IEEE Transactions on Biomedical Engineering*, *68*(5), 1557–1568.

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigne, A., Lalor, E., Meyer, B. T., Miran, S., Francart, T., & Bertrand, A. (2021). Electroencephalography-based auditory attention decoding: Toward neurosteered hearing devices. *IEEE Signal Processing Magazine*, *38*(4), 89–102.

Geravanchizadeh, M., & Roushan, H. (2021). Dynamic selective auditory attention detection using RNN and reinforcement learning. *Scientific Reports*, *11*(1), 1–11.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., & Poeppel, D. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party. *Neuron*, *77*(5), 980–991.

Griffin, A. M., Poissant, S. F., & Freyman, R. L. (2019). Speech-in-noise and quality-of-life measures in school-aged children with normal hearing and with unilateral hearing loss. *Ear and Hearing*, *40*(4), 887.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, *13*(4), 622–635.

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*(3), 583–590.

Horwitz-Martin, R. L., Quatieri, T. F., Godoy, E., & Williamson, J. R. (2016). A vocal modulation model with application to predicting depression severity. In *2016 IEEE 13th International Conference on Wearable and Implantable Body Ssensor Networks (BSN)*, IEEE, pp. 247–253.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 186.

Johari, K., den Ouden, D.-B., & Behroozmand, R. (2019). Behavioral and neural correlates of normal aging effects on motor preparatory mechanisms of speech production and limb movement. *Experimental Brain Research*, *237*(7), 1759–1772.

Larson, E., & Lee, A. K. C. (2014). Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *Neuroimage*, *84*, 681–687.

Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hämäläinen, M., & Shinn-Cunningham, B. G. (2013). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience*, *6*, 190.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PloS One*, *11*(9), e0162726.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones: Naval Health Research Center San Diego CA.

McCloy, D. R., Lau, B. K., Larson, E., Pratt, KAI, & Lee, A. drianK. C. (2017). Pupillometry shows the effort of auditory attention switching. *The Journal of the Acoustical Society of America*, *141*(4), 2440–2451.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236.

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., & Babadi, B. (2018). Real-time tracking of selective auditory attention from M/EEG: A Bayesian filtering approach. *Frontiers in Neuroscience*, *12*, 262.

Miran, S., Presacco, A., Simon, J. Z., Fu, M. C., Marcus, S. I., & Babadi, B. (2020). Dynamic estimation of auditory temporal response functions via state-space models with gaussian mixture process noise. *PLoS Computational Biology*, *16*(8), e1008172.

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., & Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research*, *377*, 260–270.

O'Sullivan, J., Chen, Z., Herrero, J., McKhann, G. M., Sheth, S. A., Mehta, A. D., & Mesgarani, N. (2017). Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *Journal of Neural Engineering*, *14*(5), 056001.

O'Sullivan, A. E., Lim, C. Y., & Lalor, E. C. (2019). Look at me when I'm talking to you: Selective attention at a multisensory cocktail party can be decoded using stimulus reconstruction and alpha power modulations. *European Journal of Neuroscience*, *50*(8), 3282–3295.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706.

Paul, B. T., Chen, J., Le, T., Lin, V., & Dimitrijevic, A. (2021). Cortical alpha oscillations in cochlear implant users reflect subjective listening effort during speech-in-noise perception. *Plos One*, *16*(7), e0254162.

Pfurtscheller, G. (2001). Functional brain imaging based on ERD/-ERS. *Vision Research*, *41*(10-11), 1257–1260.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, BWY, Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., & Naylor, G. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). *Ear and Hearing*, *37*, 5S–27S.

Popelka, G. R., & Moore, B. C. J. (2016). Future directions for hearing aid development, *Hearing aids*: Springer, pp. 323–333.

Posner, M. I., Cohen, Y., et al. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556.

Presacco, A., Simon, J. Z., & Anderson, S. (2019). Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PloS One*, *14*(3), e0213899.

Puvvada, K. C., & Simon, J. Z. (2017). Cortical representations of speech in a multitalker auditory scene. *Journal of Neuroscience*, *37*(38), 9189–9196.

Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., & Lunner, T. (2020). An exploratory study of EEG alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. *Plos One*, *15*(7), e0235782.

Senkowski, D., Saint-Amour, D., Gruber, T., & Foxe, J. J. (2008). Look who's talking: The deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *Neuroimage*, *43*(2), 379–387.

Stephen, J. M. (2019). Designing MEG experiments. In *Magnetoencephalography: From signals to dynamic cortical networks* (pp. 205–235). Springer.

Teoh, E. S., & Lalor, E. C. (2019). Eeg decoding of the target speaker in a cocktail party scenario: Considerations regarding dynamic switching of talker location. *Journal of Neural Engineering*, *16*(3), 036017.

Van Canneyt, J., Wouters, J., & Francart, T. (2021). Cortical compensation for hearing loss, but not age, in neural tracking of the fundamental frequency of the voice. bioRxiv.

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 2331216519832483.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *Journal of the Association for Research in Otolaryngology*, *19*(2), 181–191.

Weisz, N., Hartmann, T., Müller, N., & Obleser, J. (2011). Alpha rhythms in audition: Cognitive and clinical perspectives. *Frontiers in Psychology*, *2*, 73.

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*, 2331216518800869.

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

How to cite this article: Haro, S., Rao, H. M., Quatieri, T. F., & Smalt, C. J. (2022). EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. *European Journal of Neuroscience*, *55*(5), 1262–1277. https://doi.org/10.1111/ejn.15616