OPEN

Exploring N400 Predictability Effects During Sustained Speech Comprehension: From Listening-Related Fatigue to Speech Enhancement Evaluation

Cheng-Hung Hsin,[1] Chia-Ying Lee,[2,3,4] and Yu Tsao[1,5]

Objectives: This study investigated the predictability effect on the N400 as an objective measure of listening-related fatigue during speech comprehension by: (1) examining how its characteristics (amplitude, latency, and topographic distribution) changed over time under clear versus noisy conditions to assess its utility as a marker for listening-related fatigue, and (2) evaluating whether these N400 parameters could assess the effectiveness of speech enhancement systems.

Design: Two event-related potential experiments were conducted on 140 young adults (aged 20 to 30) assigned to four age-matched groups. Using a between-subjects design for listening conditions, participants comprehended spoken sentences ending in high- or low-predictability words while their brain activity was recorded using electroencephalography. Experiment 1 compared the predictability effect on the N400 in clear and noise-masked conditions, while experiment 2 examined this effect under two enhanced conditions (denoised using the Transformer- and minimum mean square error-based speech enhancement models). Electroencephalography data were divided into two blocks to analyze the changes in the predictability effect on the N400 over time, including amplitude, latency, and topographic distributions.

Results: Experiment 1 compared N400 effects across blocks under different clarity conditions. Clear speech in block 2 elicited a more anteriorly distributed N400 effect without reduction or delay compared with block 1. Noisy speech in block 2 showed a reduced, delayed, and posteriorly distributed effect compared with block 1. Experiment 2 examined N400 effects during enhanced speech processing. Transformer-enhanced speech in block 1 demonstrated significantly increased N400 effect amplitude compared to noisy speech. However, both enhancement methods showed delayed N400 effects in block 2.

Conclusions: This study suggests that temporal changes in the N400 predictability effect might serve as objective markers of sustained speech processing under different clarity conditions. During clear speech comprehension, listeners appear to maintain efficient semantic processing through additional resource recruitment over time, while noisy speech leads to reduced processing efficiency. When applied to enhanced speech, these N400 patterns reveal both the immediate benefits of speech enhancement for semantic processing and potential limitations in supporting sustained listening. These findings demonstrate the potential utility of the N400 predictability effect for understanding sustained listening demands and evaluating speech enhancement effectiveness.

Key words: Event-related potentials, Listening fatigue, N400, Predictability effect, Semantic processing, Sentence comprehension, Speech enhancement, Speech in noise.

(Ear & Hearing 2025;46;922–940)

INTRODUCTION

In everyday conversational settings, speech comprehension is performed with background noise. When noise levels increase to impair comprehension, listeners devote additional cognitive resources to maintain adequate understanding. However, sustained exertion of listening effort to process degraded speech can trigger listening fatigue (Pichora-Fuller et al. 2016), particularly in individuals with hearing impairments and those who use assistive listening devices. Listening-related fatigue has detrimental effects on quality of life, workplace productivity, and safety (Hornsby et al. 2016). It can also hinder the development of healthy hearing aid usage if the device does not adequately mitigate fatigue-inducing factors, such as noise exposure (Hong et al. 2014; Skagerstrand et al. 2014; Gygi & Ann Hall 2016). Importantly, hearing loss is associated with a higher risk of dementia (Uhlmann et al. 1989; Lin et al. 2011; Deal et al. 2017; Livingston et al. 2017; Liu & Lee 2019), while regular use of hearing aids may slow cognitive decline (Lin et al. 2023; Sarant et al. 2024). Consequently, there is an urgent need for objective measures of listening fatigue to evaluate the listening experience of individuals with identified or undiagnosed hearing issues and to assess the usability of assistive listening devices.

Although various subjective, behavioral, and physiological measures exist to quantify listening effort and fatigue, such as self-reporting, reaction time, and brain activity recordings (for existing reviews, see McGarrigle et al. 2014; Hornsby et al. 2016; Pichora-Fuller et al. 2016; Ohlenforst et al. 2017; Francis & Love 2019), neurophysiological and neuroimaging techniques provide unique insights into the allocation of cognitive resources under adverse listening conditions that increase effort and fatigue. In particular, electroencephalography (EEG) offers excellent temporal resolution on the millisecond scale, enabling the real-time tracking of neural processes involved in auditory stimulus processing. Studies have demonstrated that event-related potentials (ERPs) are reliable markers of listening effort (Ohlenforst et al. 2017). Crucially, ERPs reflect cognitive processes related to attention and language comprehension, such as P3a, P3b, and N400 (Polich 2007; Kutas & Federmeier 2011), indicating them as appropriate physiological measures for evaluating the impact of listening effort and fatigue on cognitive processing.

Listening effort refers to the deliberate allocation of cognitive resources to overcome challenges in achieving listening-oriented

[1]Biomedical Acoustic Signal Processing Lab, Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan; [2]Brain and Language Laboratory, Institute of Linguistics, Academia Sinica, Taipei, Taiwan; [3]Institute of Cognitive Neuroscience, National Central University, Taoyuan, Taiwan; [4]Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei, Taiwan; and [5]Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan.

<zdoi; 10.1097/AUD.0000000000001635>

goals (Pichora-Fuller et al. 2016; Francis & Love 2019). Related ERP studies have used dual-task and oddball paradigms to investigate how listening effort affects P3a and P3b responses under various auditory conditions in different age and hearing ability groups (Bertoli & Bodmer 2014, 2016; Kestens et al. 2023). P3a, reflecting frontal attention mechanisms, is triggered by novel or unexpected stimuli and indicates attention reorientation to task-irrelevant distractors (Polich 2007). Using speech-perception-in-noise tests, Bertoli and Bodmer (2016) observed an age-related decline in P3a amplitude to task-irrelevant novel sounds, indicating reduced attention reorientation capacity in older adults when a demanding primary task taxes cognitive resources. P3b, generated in temporal-parietal regions, reflects attentional resource allocation and memory updating during task-relevant stimulus processing (Polich 2007). In a word discrimination study, Kestens et al. (2023) found that noise increased P3b amplitude and latency, reflecting greater listening effort, while aging decreased these responses. Although older adults exhibited reduced P3b amplitudes and prolonged latencies compared with younger adults, no interaction emerged between age and noise conditions, indicating that both age groups showed similar patterns of increased P3b measures under noisy conditions. These findings show that P3b is sensitive to both cognitive demands under adverse listening conditions and age-related changes in auditory processing.

Few studies have used ERP measures to assess listening fatigue (Key et al. 2017; Moore et al. 2017; Gustafson et al. 2018). Listening fatigue results from the sustained application of effort during challenging listening tasks (Pichora-Fuller et al. 2016). Therefore, the duration of a listening task is a crucial factor when measuring fatigue (McGarrigle et al. 2014; Hornsby et al. 2016). Hornsby et al. (2016) proposed that the impact of fatigue on cognitive processing could be measured using a mentally demanding task allowing the continuous monitoring of cognitive performance over an extended period (termed the direct approach). Alternatively, cognitive capabilities can be evaluated before and after a prolonged period to assess fatigue-related effects (indirect approach). For example, Key et al. (2017) used an indirect approach in which normal-hearing children underwent extensive speech-processing tasks across three-hour sessions designed to measure changes in fatigue. The P300 responses were recorded before and after these sessions using the oddball paradigm with a syllable discrimination task. This study found that parietal P300 (reflecting P3b) amplitude decreased after the sessions, indicating a reduced ability to discriminate syllables due to fatigue. In contrast, Moore et al. (2017) used a direct approach utilizing a 50-min auditory choice task divided into two 25-min blocks to explore mental fatigue resulting from prolonged auditory processing. The task required participants to discern the sequences of three pure tones and determine whether they predominantly heard high- or low-frequency tones. The study showed a significant reduction in the N1 amplitude during the second block compared with the first, reflecting decreased sensory arousal and attentional engagement over time. These findings underscore the potential of the N1 and P3b components as neural markers for assessing listening fatigue.

Among the ERP measures explored, one remaining challenge is that these paradigms do not utilize complex linguistic materials (e.g., sentences) to assess how listening fatigue affects higher-order language comprehension, a critical aspect of real-world communication. The ERP N400, a well-established neural signature of semantic processing (Kutas & Hillyard 1980, 1984), can address this challenge. The N400 emerges around 200 to 300 msec, peaking around 400 msec over the centroparietal regions of the scalp during reading tasks. In listening tasks, the N400 often exhibits an earlier onset and a longer duration, with a broader scalp distribution extending anteriorly toward frontal regions (Kutas & Van Petten 1994; Van Petten & Luka 2006). The N400 is particularly sensitive to a word's cloze probability in context (Kutas & Federmeier 2011). The cloze probability (operationalized as predictability) of a sentence-final word, where the N400 is typically measured, is defined as the percentage of participants who provide that word as the most probable sentence completion (Taylor 1953; Kutas & Federmeier 2000). N400 amplitudes exhibit an inverse relationship with word predictability, such that less predictable words elicit enhanced N400 amplitudes relative to predictable words within a given context. The N400 predictability effect reflects facilitated word access and semantic integration, as listeners can pre-activate lexical items or features to promote comprehension by utilizing contextual information. Thus, predictable words are processed more efficiently, whereas unpredictable ones require greater cognitive effort (Kutas & Federmeier 2011).

Research has extensively utilized the N400 predictability effect to investigate the neural mechanisms underlying speech comprehension (Holcomb & Neville 1991; van den Brink et al. 2006; Boudewyn et al. 2015). A growing body of research on the N400 has examined how listening conditions affect speech comprehension. The N400 predictability effect is sensitive to various adverse conditions across the speech chain (Mattys et al. 2012; Peelle 2018). Specifically, this N400 effect is reduced and delayed as speech clarity decreases due to compromised speaker, environmental, or listener conditions, such as accented speech (Romero-Rivas et al. 2016; Grey & van Hell 2017; Holt et al. 2018; Wambacq et al. 2023), noisy speech (Obleser & Kotz 2011; Strauß et al. 2013; Silcox & Payne 2021; Hsin et al. 2023), and hearing impairment (Hahne et al. 2012; Burkhardt et al. 2022). Research has suggested that degraded speech signals narrow listeners' lexical expectations due to constrained context availability, reducing the N400 predictability effect as comprehension becomes effortful (Strauß et al. 2013).

The topographic distribution of N400 effects provides valuable insight into the cognitive resource allocation during language comprehension. In healthy adults, the distribution and timing of N400 effects can vary based on cognitive demands and listening conditions. During spoken sentence comprehension, Li and Ren (2012) demonstrated that heightened word accentuation led to more anteriorly distributed N400 congruency effects with earlier onsets compared with normal accentuation, suggesting increased allocation of attentional resources for deeper processing and earlier detection of semantic anomalies. Similarly, Boudewyn et al. (2013) found that listeners with higher working memory (WM) capacity showed right-frontal N400 congruency effects during discourse comprehension, contrasting with the posterior distribution in lower WM listeners. This pattern indicates that additional cognitive resources may be recruited to maintain and integrate context information when processing capacity allows.

Studies examining speech comprehension under various conditions consistently report differences in N400 topography between clear and adverse listening conditions. Clear speech typically elicits more anteriorly distributed N400 effects

compared with adverse conditions, such as accented or noisy speech (Strauß et al. 2013; Grey & van Hell 2017; Hsin et al. 2023; Wambacq et al. 2023). Strauß et al. (2013) suggested this pattern might reflect different neural source configurations, with clear speech potentially engaging broader neural resources. Supporting this interpretation, functional magnetic resonance imaging (fMRI) studies show that left inferior frontal gyrus (IFG) activity increases with speech intelligibility, indicating the IFG's role in modulating neural resources to facilitate comprehension as speech clarity improves (Obleser & Kotz 2010).

The importance of distinct resource recruitment is further highlighted by lesion studies, where individuals with brain injuries show anteriorly distributed N400 effects during effortful comprehension, suggesting additional or compensatory mechanisms for lexical processing (Chang et al. 2016; Ledwidge et al. 2022). Clinical studies have shown that these neural patterns can be modified through intervention—intensive speech therapy can shift a right-frontal N400 effect to a left-central distribution during speech-processing tasks (Wilson et al. 2012b), demonstrating the plasticity of resource allocation patterns in response to therapeutic intervention.

Studies of cognitive fatigue in nonlinguistic tasks provide additional evidence linking sustained effort with frontal activity. ERP and positron emission tomography studies have shown shifts in activation patterns toward frontal regions during prolonged cognitive tasks (Wang et al. 2016; Sabeti et al. 2018), with increased medial orbitofrontal activation correlating with subjective fatigue ratings (Tajima et al. 2010). However, it remains unclear whether the N400 effect exhibits similar topographic shifts over time during speech comprehension, which could provide insights into how cognitive resources are reallocated when listeners experience fatigue under different clarity conditions.

In practice, word predictability has been applied to develop the speech perception in noise test, a standardized clinical assessment tool (Kalikow et al. 1977; Bilger et al. 1984; Wilson et al. 2012a). Behavioral studies have repeatedly demonstrated predictability effects during spoken sentence comprehension under both clear and adverse conditions (Pichora-Fuller et al. 1995; Sommers & Danielson 1999; Sheldon et al. 2008). Notably, the N400 predictability effect does not require overt behavioral responses to be elicited (Kutas & Federmeier 2011). Jamison et al. (2016) used a passive listening task to examine the feasibility of using the N400 based on word predictability to estimate listeners' speech-in-noise thresholds. Therefore, N400 measures based on lexical predictions are particularly suitable in populations that are unable to fully cooperate in hearing tests (e.g., patients). The N400 predictability effect demonstrated its potential as a measure of listening-related fatigue.

This study aimed to evaluate the utility of the predictability effect on the N400 (hereinafter referred to as "the N400 effect") as an objective marker of listening-related fatigue during speech comprehension. Across two experiments, we examined the temporal modulation of the N400 effect, encompassing amplitude, latency, and topographic distribution changes. Participants were asked to comprehend spoken sentences ending with high- or low-predictability words under clear, noisy, and enhanced conditions. A mixed factorial design was used, with the listening condition as a between-subjects factor, and predictability (high versus low) and time (first versus second block) as within-subjects factors. The between-subjects manipulation of listening condition was chosen to isolate fatigue effects specific to each clarity condition, avoiding potential carryover effects between conditions. This design allowed us to discern the influence of listening fatigue on comprehension that can be attributed solely to specific clarity conditions, enabling us to examine how different speech clarity conditions differentially induce listening fatigue over time. In experiment 1, two age-matched groups comprehended clear or noise-masked sentences. In experiment 2, two additional age-matched groups comprehended denoised versions of the noise-masked sentences from experiment 1, which were enhanced using a machine-learning-based Transformer or conventional statistical model-based minimum mean square error technique (hereinafter referred to as Transformer and MMSE, respectively).

Few studies have used ERP measures to assess the effectiveness of speech enhancement systems. Using mismatch negativity (MMN), an ERP component that reflects preattentive detection of auditory changes, Yu et al. (2018) showed that noise suppression algorithms can benefit early auditory processing. Specifically, in their oddball paradigm with vowel stimuli, enhanced speech elicited larger MMN amplitudes with shorter latencies compared with unprocessed noisy speech, indicating more efficient automatic detection of acoustic changes (see also Yu et al. 2019). While these findings demonstrate speech enhancement systems' effectiveness at the early auditory processing stage, their impact on higher-order language comprehension and listening-related fatigue remains unexplored. Currently, no established physiological measures exist to objectively evaluate how speech enhancement systems might mitigate listening-related fatigue. Therefore, experiment 2 was designed to investigate the utility of the N400 predictability effect for assessing the effectiveness of speech enhancement systems and the associated user experience.

As cognitive resources are inherently limited (Kahneman 1973; Pichora-Fuller et al. 2016), listening fatigue is expected to reduce the attentional capacity for semantic processing over time. N400 effects can reflect participants' attentional levels during listening tasks (Cruse et al. 2014; Erlbeck et al. 2014; Kemp et al. 2019). Therefore, we compared the N400 effects between the first and second experimental blocks to investigate whether the temporal changes of this effect can reflect listening-related fatigue during speech comprehension over time (i.e., the direct approach). In experiment 1, we developed specific hypotheses about how the N400 predictability effect might change over time under different listening conditions. For clear speech, we expected preserved semantic processing would be reflected in maintained N400 effect amplitudes across blocks, with no significant latency delays due to efficient processing. However, we hypothesized that sustained listening might lead to fatigue-related resource recruitment, manifesting as broader scalp distributions in later blocks. For noisy speech, we predicted progressive deterioration in semantic processing efficiency. Specifically, we hypothesized that cognitive resource depletion would result in reduced N400 effect amplitudes and increasingly delayed latencies across blocks. In contrast to clear speech, we expected the effect to remain primarily centroparietal across blocks, as the demands of noise processing might prevent the recruitment of additional resources.

In experiment 2, we examined how speech enhancement might affect these temporal patterns. We hypothesized that both enhancement methods would show N400 patterns more similar to clear than noisy speech, with increased amplitudes, earlier onsets, and broader topographic distributions. Given its

advanced technology, we expected the Transformer-based model to demonstrate more significant N400 effects than the MMSE model. The critical test would come in later blocks. If speech enhancement systems effectively reduce listening fatigue, these improved patterns should be maintained over time. Alternatively, if enhancement introduces its own processing demands, we might observe progressive changes in the N400 effect characteristics similar to those seen in noise. Evidence on how time and acoustic conditions influence the N400 effect will illustrate the utility of this multi-dimensional neural marker in measuring listening fatigue for potential real-world applications.

EXPERIMENT 1: MEASURING LISTENING FATIGUE DURING COMPREHENSION OF CLEAR AND NOISY SPEECH

Materials and Methods

Participants • Experiment 1 recruited two groups of age-matched healthy young adults to investigate listening fatigue during spoken sentence comprehension under clear and noisy conditions. Thirty-five adults participated in the clear condition (14 males, aged 20 to 30 years, mean = 24.09 ± 3.04), and a different group of 35 adults participated in the noisy condition (14 males, aged 20 to 30 years, mean = 24.23 ± 3.25). All were right-handed native Mandarin Chinese speakers (mean years of education: clear group = 16.06 ± 1.59; noisy group = 15.8 ± 1.59), with normal or corrected-to-normal vision and no reported neurological/psychiatric disorders or brain damage. Pure-tone audiometry (PTA) confirmed normal-hearing thresholds below 25 dB averaged across 500, 1000, 2000, and 4000 Hz in each participant's better ear. All recruitment and experimental procedures, including informed consent and data privacy protocols, complied with the ethical human research conduct regulated by the local institutional review board (IRB).

Stimuli • This study adopted stimuli from Hsin et al. (2023), which included 128 Mandarin Chinese sentences ending in high- or low-predictability final words. Sixty-four highly constraining sentences rendered the final words highly predictable, whereas the remaining 64 weakly constraining sentences rendered the final words less predictable. These sentences are semantically plausible and syntactically simple. Each sentence contained 9 to 16 characters and ended with a disyllabic final word. Final words were matched for word frequency, contextual diversity, semantic diversity, visual complexity, and orthographic neighborhood size (Hsin et al. 2023). Additional analyses confirmed that the final words were also matched for phonological properties, including phonological neighborhood size (t = 0.22, p = 0.83), uniqueness point (U = 2048.0, Z = 0.0, p = 1.0), and phoneme distribution [initials: $\chi^2(20, N = 214) = 1.75$, $p = 1.00$; finals: $\chi^2(14, N = 342) = 2.90$, $p = 0.99$; lexical tones: $\chi^2(3, N = 255) = 2.56$, $p = 0.46$].

Auditory stimuli were created using Audacity software (version 2.1.3). A female native Mandarin speaker recorded the 128 sentences at 16 bits/44.1 kHz in a quiet room, maintaining natural speed, intonation, and prosody (mean sentence length = 4.1 ± 0.61 sec; mean target word length = 0.78 ± 0.1 sec). The noisy condition used babble noise derived from recordings of indoor social gatherings (BBC Sound Effects library, https://sound-effects.bbcrewind.co.uk/). These recordings feature multiple overlapping voices without intelligible speech content, characteristic of cocktail party environments. Each sentence was masked by a different babble snippet, spanning from sentence onset to offset, to avoid potential learning effects from repeated noise patterns. On the basis of environmental studies, the masking used a lower-bound signal to noise ratio (SNR) of +2 dB (Smeds et al. 2015; Wu et al. 2018). The stimuli were normalized to a 65 dB SPL, approximating the daily conversation level.

Procedure • Participants were individually seated approximately 90 cm in front of a liquid-crystal display in a sound-attenuated and electrically shielded room. The experiments were presented using MATLAB (R2014b) and Psychtoolbox 3. Auditory stimuli were presented binaurally through two loudspeakers positioned 90 cm in front of each participant. Before EEG recording, participants were instructed to minimize muscle and ocular artifacts during stimulus presentation. Subsequently, the participants underwent six practice trials to familiarize themselves with the experimental procedures. The researchers ensured that each participant could hear and read the stimuli clearly, using a sound level meter controlling the maximum intensity at 65 dB SPL. The formal session comprised 128 randomized trials delivered across four segments. Each segment lasted approximately 7 to 8 min, with a 90-sec break between segments. Given that behavioral tasks can influence N400 responses through strategy development and task environment effects (Brothers et al. 2017; Kemp et al. 2019; Lai et al. 2023; for a review, see Federmeier 2021), the experiment included questions on only 25% of trials, requiring either true/false or word-choice responses. These occasional questions were included to ensure participants remained attentive during stimulus presentation while maintaining a lower cognitive load, as frequent responses could incur additional mental fatigue through task-related memory retrieval and decision-making. This minimal response design allowed us to focus on neural changes associated explicitly with listening comprehension.

Each trial began with a 500 msec fixation cross to prepare participants for the upcoming stimulus. A spoken sentence was then presented once during the fixation. Participants attended to the stimulus for listening comprehension (2500 to 5000 msec). The fixation remained on the screen for an additional 2000 msec after the stimulus. In 25% of the trials, a true/false or forced word-choice question regarding the preceding sentence appeared onscreen. For example, after hearing *Never sell your soul to the devil*, participants might see *Do you sell your soul to the devil?—True/False* or *Which word did you hear?—devil/wire*. The participants responded using a keyboard, and the computer recorded their behavioral accuracy. For trials without questions (75%), participants were prompted to press a button to advance to the subsequent trial. All trials included a 1.6-sec intertrial interval, which began after the participant responded to a question or after their button press to advance. The questions were counterbalanced between those targeting the final word and those addressing overall sentence meaning. To minimize EEG artifacts, participants were instructed to respond after the question onset. The experiment continued after the participant responded or pressed a button to advance.

EEG/ERP Acquisition • EEG/ERPs were recorded from 64 sintered Ag/AgCl electrodes (QuickCap, Neuromedical Supplies) referenced to the scalp vertex between Cz and CPz. The ground electrode was positioned anterior to Fz. Signals were

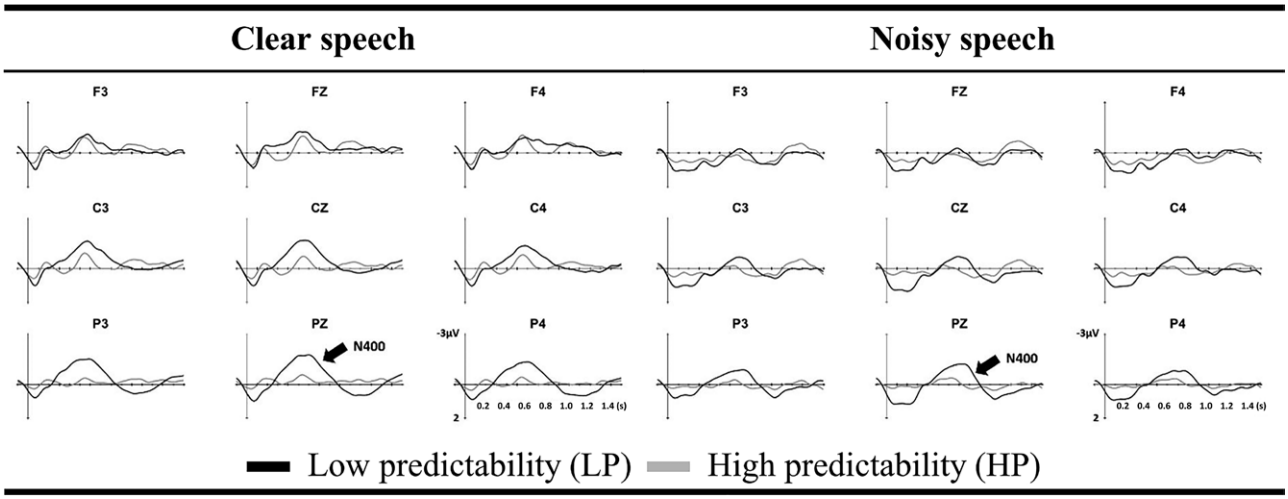

Fig. 1. Averaged ERMs of the low- (black) and high- (gray) predictability words under the clear (left) and noisy (right) conditions. ERMs indicate event-related modes.

continuously digitized at 1000 Hz using SynAmps2 (Neuroscan Inc.), with a low-pass filter at 400 Hz for offline analysis, and re-referenced to the averaged mastoids (M1/M2). Vertical eye movements and blinks were monitored using supraorbital and infraorbital electrodes on the left eye. Horizontal movements were monitored using outer canthi electrodes. Impedances were maintained below 5 kΩ.

EEG/ERP Preprocessing • EEG data were preprocessed using FieldTrip (Oostenveld et al. 2011). Epochs were segmented and time-locked to the target word onset from −100 msec prestimulus onset to 1500 msec poststimulus onset. The prestimulus interval was used for baseline correction. Data were bandpass filtered within 0.1 and 30 Hz across all channels. Artifact correction was applied to epochs containing eye movements. Trials with voltage variations larger than 100 μV were rejected. The trial rejection rates for clear and noisy conditions were 3.59 and 4.57%, respectively. The remaining trials underwent ensemble empirical mode decomposition (EEMD) analysis (Huang et al. 1998; Wu & Huang 2009).

EEMD Analysis • EEMD is a data-driven approach for the analysis of nonlinear and nonstationary time-frequency data, such as EEG/ERP. This adaptive algorithm optimizes SNRs to improve ERP latency and amplitude estimates (for a review of existing studies, see Sweeney-Reed et al. 2018). This study followed the EEMD analysis procedure described by Hsu et al. (2016) to generate intrinsic mode functions (IMFs) representing time-frequency local event properties. Gaussian white noise was added to the signal of each EEG trial for each channel to compute the EEMD. The Gaussian noise amplitude was set at 10% of the standard deviation (SD) of the EEG signal. The noise-assisted signal was decomposed into eight IMFs and one residual trend, sifting through 10 repetitions with different white noises to produce 40 ensembles of the corresponding IMFs. The resultant IMFs were obtained by averaging all ensembles of each IMF. Hilbert spectrum analysis was applied to evaluate the time-frequency spectra of the IMFs. The results revealed that the central frequencies of IMF5 to IMF8 were 11, 5, 2, and 1 Hz. On the basis of previous N400 studies on speech processing (Chen et al. 2016; Chao et al. 2019; Hsin et al. 2023), IMFs 6, 7, and 8 were summed and averaged over all trials per condition in each channel to yield event-related modes (ERMs).

Statistical Analysis • Figure 1 shows the grand-averaged ERMs elicited by low- and high-predictability words, with the N400 component visible from 200 to 900 msec in the centroparietal and frontocentral regions. Experiment 1 aimed to determine whether listening fatigue modulates predictability effects on N400 during clear and noisy speech comprehension. First, we examined the overall predictability effects for each clarity condition. The subsequent fatigue analysis divided the data into two blocks corresponding to the first and second halves of the experiment (Fig. 2). To ensure consistent data quality across blocks, we examined trial retention: the clear condition retained an average of 61 (block 1) and 62 (block 2) trials out of 64, while the noisy condition retained an average of 61 trials in both blocks. Cluster-based permutation tests (CBPT; Maris & Oostenveld 2007) were conducted to characterize the temporal dynamics and topographic distributions of the predictability effects. Linear mixed models (LMMs) analysis (Baayen et al. 2008) was applied to examine whether clarity and block modulated the N400 effect and validate the CBPT findings.

Temporal and Spatial CBPT • Given the exploratory nature of examining N400 modulation by listening fatigue over time, we used CBPT for the primary analyses. This approach was particularly suitable because: (1) as one of the first studies investigating N400 and listening fatigue, we lacked a priori assumptions about how block effects under different clarity conditions might manifest in spatial and temporal domains; (2) the effects of different SNR levels, noise types, and speech enhancement on N400's temporal and spatial characteristics were unknown, especially as the auditory N400 tends to exhibit longer latencies and broader scalp distributions compared with the visual N400, making it challenging to predefine time windows and regions of interest (ROIs) using traditional approaches; and (3) CBPT effectively handles the multiple comparison problem in high-dimensional EEG data while maintaining sensitivity to detect effects across both spatial and temporal domains (Maris & Oostenveld 2007; Sassenhagen & Draschkow 2019).

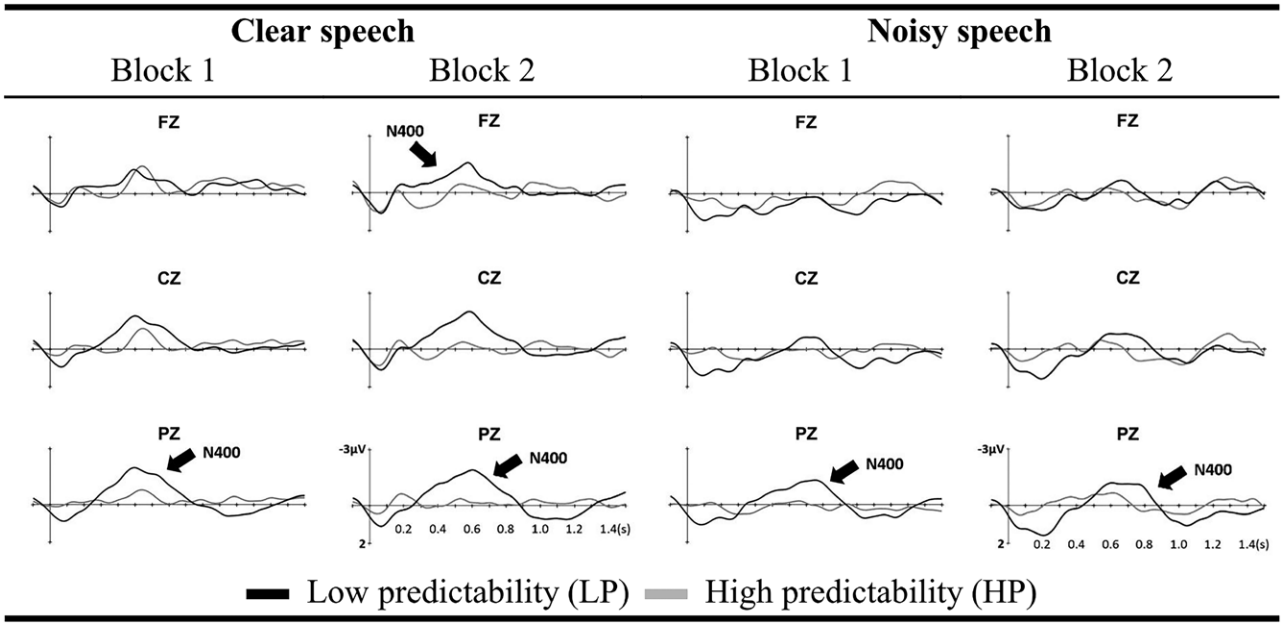

Fig. 2. Averaged ERMs of the low- (black) and high- (gray) predictability words in blocks 1 and 2 under the clear (left) and noisy (right) conditions. ERMs indicate event-related modes.

CBPT is a nonparametric method considering correlated neural activities at adjacent time points and electrodes when estimating effects. This study used FieldTrip to perform CBPT to examine the predictability effects on the N400 for clear and noisy conditions. Epochs from 200 to 900 msec were divided into 14 successive 50-msec time windows per condition and electrode. The tests were conducted on the mean amplitudes of each time window for each electrode using the following steps: first, for predictability effects in each clarity condition, the algorithm performed a simple dependent-samples t test at each time point and each electrode, identifying points and electrodes exceeding significance ($\alpha = 0.05$) to form clusters. Temporal and spatial clusters were formed when the t statistics of two consecutive time points and two neighboring electrodes exceeded significance. Cluster mass was calculated for each cluster by summing all the individual t statistics within that cluster. Subsequently, the algorithm randomly shuffles the condition codes to create a null distribution of maximum cluster masses over 1000 permutations, identifying clusters at each permutation and finding the maximum cluster mass. Cluster correction was used to control the familywise error rate for multiple comparisons. Last, the observed cluster masses were compared against the null distribution, and only the cluster masses in the highest or lowest 2.5th percentile were considered significant.

LMM Analysis • LMMs were conducted to examine whether clarity and block modulated the N400 predictability effect in fixed time windows and ROIs, and to validate the CBPT findings. The analysis was implemented using the lme4 package (version 1.1.23) (Bates et al. 2015) in R (version 4.0.0; R Core Team 2021). Using single-trial data, the LMMs calculated the ERP mean amplitude differences between low- and high-predictability sentences for clear and noisy conditions across blocks in the temporal windows of interest over ROIs. Temporal windows and ROIs were defined based on the CBPT results on the main effects of predictability. Temporal windows of interest were defined from 300 to 850 msec for the N400. Two ROIs were defined: anterior (Fz, F1, F2, F3, F4, FCz, FC1, FC2, FC3, FC4, Cz, C1, C2, C3, and C4), and posterior (CPz, CP1, CP2, CP3, CP4, Pz, P1, P2, P3, P4, POz, PO3, PO4, PO5, and PO6). Because LMMs simultaneously model participant and item variance as random effects, models were estimated by including predictability (high/low), block (first/second), clarity (clear/noisy), and the interaction between these effects in the ROIs as fixed effects, and the participants and items as crossed random effects. The estimated coefficient (β), standard error (SE), t value, and p value for fixed effects are listed in Table 1.

Results

Behavioral Performance • Participants in the clear and noisy conditions achieved overall mean accuracies of 99.24% (SD = 2.01, range = 92.86 to 100%) and 98.91% (SD = 2.44, range = 91.67 to 100%), respectively. A two-way repeated-measures analysis of variance was conducted to analyze the effects of predictability and clarity on accuracy. The results showed a significant interaction between predictability and clarity ($F_{1, 68} = 10.47$, $p = 0.002$).

The simple main effects analysis revealed two main findings. First, in the noisy condition, accuracy was higher for high-predictability sentences (mean = 99.7%, SE = 0.37) than for low-predictability sentences (mean = 98.1%, SE = 0.37). This difference was significant ($\beta = 1.513$, SE = 0.514, t ratio = 2.942, $p = 0.004$). No such predictability effect emerged in the clear condition, where high-predictability sentences (mean = 98.8%, SE = 0.37) and low-predictability sentences (mean = 99.7%, SE = 0.37) showed comparable accuracy ($\beta = -0.841$, SE = 0.514, t ratio = −1.634, $p = 0.106$).

Second, for low-predictability sentences, accuracy was higher in the clear condition (mean = 99.7%, SE = 0.37) than in the noisy condition (mean = 98.1%, SE = 0.37). This clarity effect was significant ($\beta = 1.513$, SE = 0.517, t ratio = 2.927, $p = 0.004$). For high-predictability sentences, no significant difference was found between clear (mean = 98.8%, SE = 0.37)

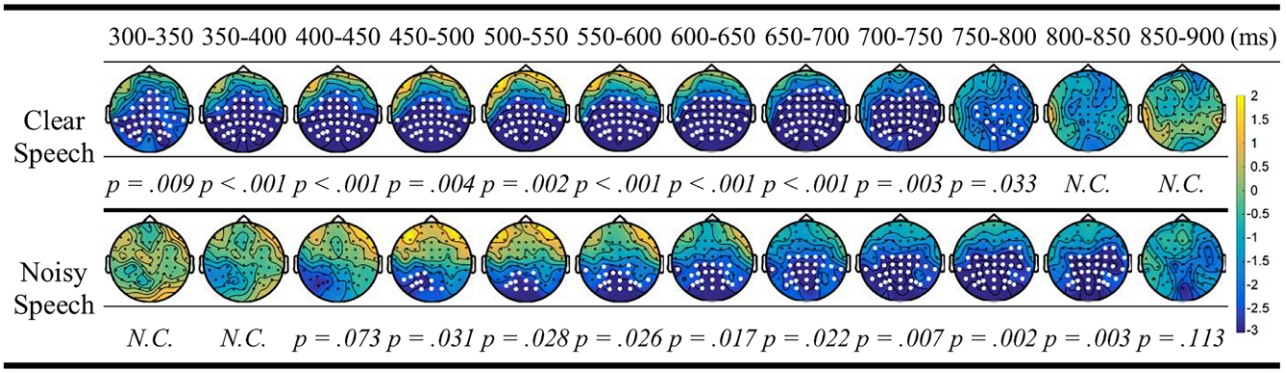

Fig. 3. Topographic maps for the predictability effect on the N400 under the clear and noisy conditions in a moving time window analysis. White dots represent electrodes with significant differences in the contrasts. N.C. indicates no clusters.

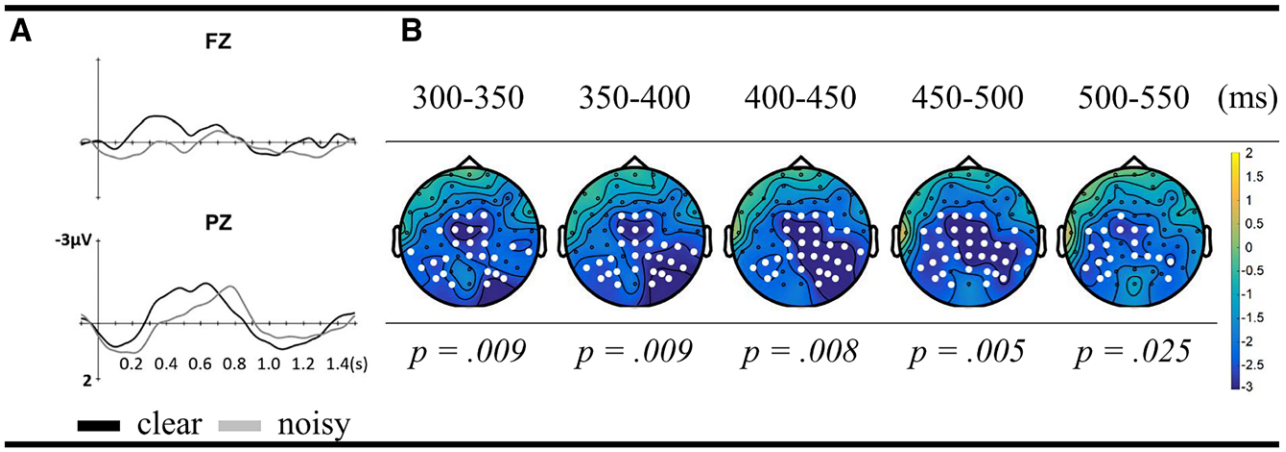

Fig. 4. Predictability effect on the N400 across clear and noisy conditions, illustrated through waveform differences and topographic distributions. A, Difference waves of the clear (black) and noisy (gray) conditions. B, Topographic maps for the predictability effect based on difference waves (clear—noisy) in a moving time window analysis. White dots represent electrodes with significant differences in the contrast.

and noisy conditions (mean = 99.7%, SE = 0.37; $\beta = -0.841$, SE = 0.517, t ratio = −1.626, $p = 0.106$).

There were no significant main effects for predictability ($F_{1, 68} = 0.86$, $p = 0.358$) and clarity ($F_{1, 68} = 0.84$, $p = 0.363$).

Cluster-Based Permutation Tests • The predictability effect (difference between low- and high-predictability conditions) manifested differently under clear and noisy conditions (Fig. 3). Under clear speech, significant negative clusters emerged in centroparietal and frontocentral regions from 300 to 800 msec ($ps < 0.05$). Under noisy speech, significant negative clusters appeared in centroparietal regions from 450 to 850 msec ($ps < 0.05$). To test the interaction between predictability and clarity, we conducted CBPT on difference waves (clear minus noisy), which revealed significant clusters in centroparietal and frontocentral regions from 300 to 550 msec ($ps < 0.05$) (Fig. 4).

The temporal dynamics of these effects varied across experimental blocks (Fig. 5). In the clear condition, block 1 showed significant negative clusters primarily in centroparietal regions from 350 to 750 msec ($ps < 0.05$), while block 2 exhibited clusters in centroparietal and frontocentral regions with an earlier onset and later offset from 300 to 850 msec ($ps < 0.01$).

In the noisy condition, block 1 showed negative clusters in left-parietal regions from 400 to 450 msec ($p = 0.043$) followed by clusters in centroparietal regions from 700 to 850 msec ($ps < 0.05$). Block 2 showed negative clusters in centroparietal and frontocentral regions from 500 to 850 msec ($ps < 0.05$), exhibiting a later onset but more consistent manifestation across time windows compared with block 1.

Compared with the clear condition, the noisy condition elicited reduced and delayed N400 effects in both blocks. To test this interaction between clarity conditions across blocks, we analyzed difference waves (Fig. 6). Block 1 showed no significant clusters ($ps > 0.05$), while block 2 revealed significant clusters in centroparietal and frontocentral regions from 200 to 650 msec ($ps < 0.05$).

LMM Analysis • Table 1 summarizes the LMM analysis examining predictability, clarity, and block effects on N400s in anterior and posterior ROIs.

Three-Way Interactions •

Predictability* × *Clarity* × *Block. The analysis revealed a significant three-way interaction between predictability, clarity, and block in both anterior ($\beta = 0.4537$, SE = 0.1293, $p < 0.01$) and posterior ROIs ($\beta = 0.5486$, SE = 0.1194, $p < 0.01$).

For the anterior ROI, this three-way interaction indicated that the predictability by clarity interaction differed between blocks. Only block 2 in the clear condition showed a significant predictability effect ($\beta = -0.9125$, SE = 0.1787, z ratio = −5.106, $p < 0.001$), with low-predictability sentences

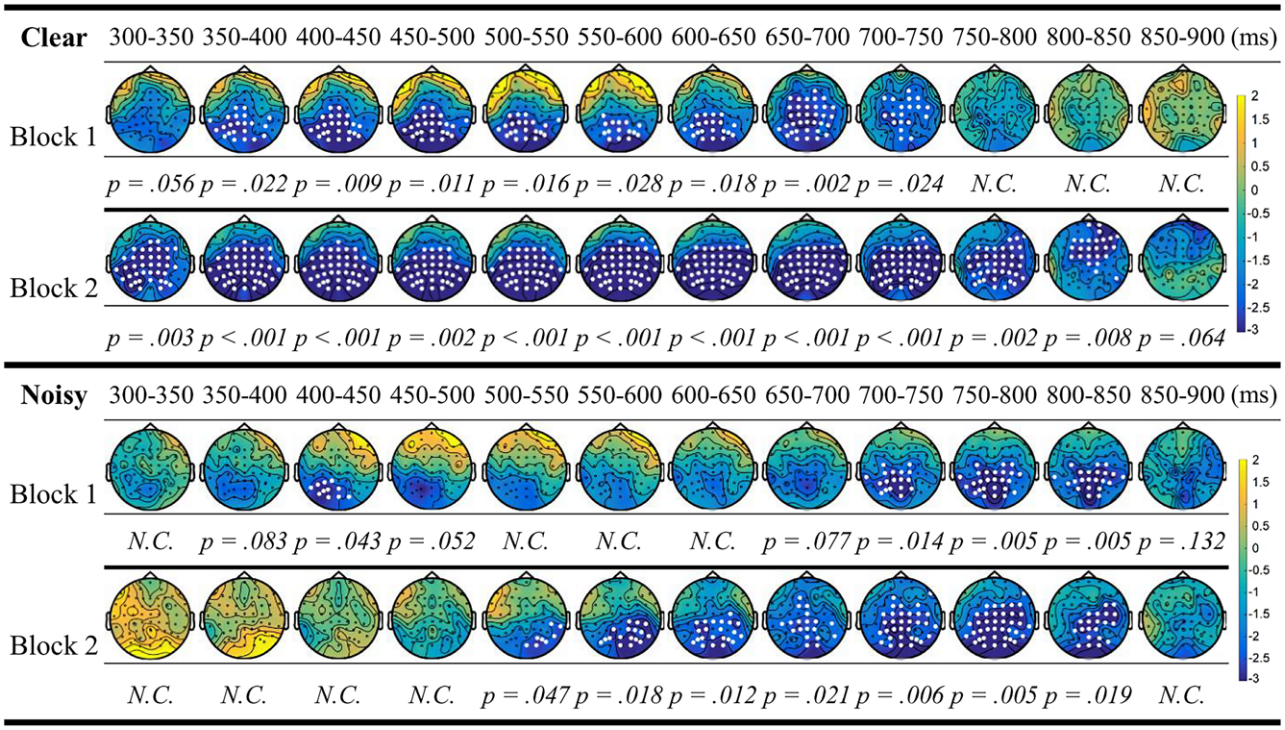

Fig. 5. Topographic maps for the predictability effect on the N400 in blocks 1 and 2 under clear (upper panel) and noisy (lower panel) conditions in a moving time window analysis. White dots represent electrodes with significant differences in the contrasts. N.C. indicates no clusters.

eliciting more negative N400s than high-predictability sentences. No significant predictability effects emerged in block 2 under noisy condition or in block 1 under either clarity condition (*ps* > 0.05).

For the posterior ROI, the three-way interaction revealed that while the predictability effect was present across conditions, its magnitude varied. The effect was largest in block 2 under clear condition ($\beta = -1.0660$, SE = 0.1479, *z* ratio = −7.207, $p < 0.001$), and smaller but significant in block 1 under both clear ($\beta = -0.7099$, SE = 0.1481, *z* ratio = −4.792, $p < 0.001$) and noisy conditions ($\beta = -0.5896$, SE = 0.1482, *z* ratio = −3.979, $p < 0.001$). The effect was not significant in block 2 under noisy condition ($p = 0.064$).

Two-Way Interactions •

Predictability* × *Clarity. Significant interactions emerged in both anterior ($\beta = 0.5544$, SE = 0.0642, $p < 0.001$) and posterior ROIs ($\beta = 0.3945$, SE = 0.0593, $p < 0.001$). In the anterior ROI, the predictability effect was significant only in the clear condition ($\beta = -0.666$, SE = 0.172, *z* ratio = −3.875, $p < 0.001$; noisy: $p = 0.847$). In the posterior ROI, the predictability effect was significant but smaller under noisy condition ($\beta = -0.493$, SE = 0.141, *z* ratio = −3.494, $p = 0.002$) compared with clear condition ($\beta = -0.888$, SE = 0.141, *z* ratio = −6.299, $p < 0.001$). In both regions, when effects were significant, low-predictability sentences yielded greater negative N400s than high-predictability sentences, with the effect being larger in the posterior region under clear condition.

Predictability* × *Block. The interaction was significant in the anterior ROI ($\beta = -0.2653$, SE = 0.0732, $p < 0.001$) but not in the posterior ROI ($p = 0.225$). In the anterior ROI, the predictability effect emerged only in block 2 ($\beta = -0.522$, SE = 0.173, *z* ratio = −3.018, $p = 0.009$; block 1: $p = 0.381$). In the posterior ROI, the predictability effect was significant in both blocks but was larger in block 2 ($\beta = -0.732$, SE = 0.142, *z* ratio = −5.153, $p < 0.001$) than in block 1 ($\beta = -0.650$, SE = 0.142, *z* ratio = −4.574, $p < 0.001$). In both regions, when effects were significant, low-predictability sentences yielded greater negative N400s than high-predictability sentences, with the effect being larger in the posterior region during block 2.

Clarity* × *Block. Significant interactions manifested in both anterior ($\beta = -0.4746$, SE = 0.0643, $p < 0.001$) and posterior ROIs ($\beta = -0.1420$, SE = 0.0594, $p = 0.017$). In the anterior ROI, the clarity effect was smaller in block 2 ($\beta = 0.403$, SE = 0.150, *z* ratio = 2.695, $p = 0.024$) than in block 1 ($\beta = 0.878$, SE = 0.150, *z* ratio = 5.867, $p < 0.001$). In the posterior ROI, the clarity effect was significant only in block 1 ($\beta = 0.376$, SE = 0.129, *z* ratio = 2.923, $p = 0.012$; block 2: $p = 0.211$). In both regions, when effects were significant, clear speech yielded greater negative N400s than noisy speech, with the effect being larger in the anterior region during block 1.

Main effects • Significant main effects of predictability emerged in both anterior ($\beta = -0.3892$, SE = 0.1691, $p = 0.023$) and posterior ROIs ($\beta = -0.6906$, SE = 0.1379, $p < 0.001$), with low-predictability sentences eliciting greater N400 negativity than high-predictability sentences. Similarly, clarity showed significant main effects in both anterior ($\beta = 0.6404$, SE = 0.1461, $p < 0.001$) and posterior ROIs ($\beta = 0.3050$, SE = 0.1251, $p = 0.017$), with clear speech eliciting greater N400 negativity than noisy speech. A main effect of block appeared only in the anterior ROI ($\beta = -0.0912$, SE = 0.0364, $p = 0.012$; posterior: $p = 0.366$), where block 2 elicited greater N400 negativity than block 1.

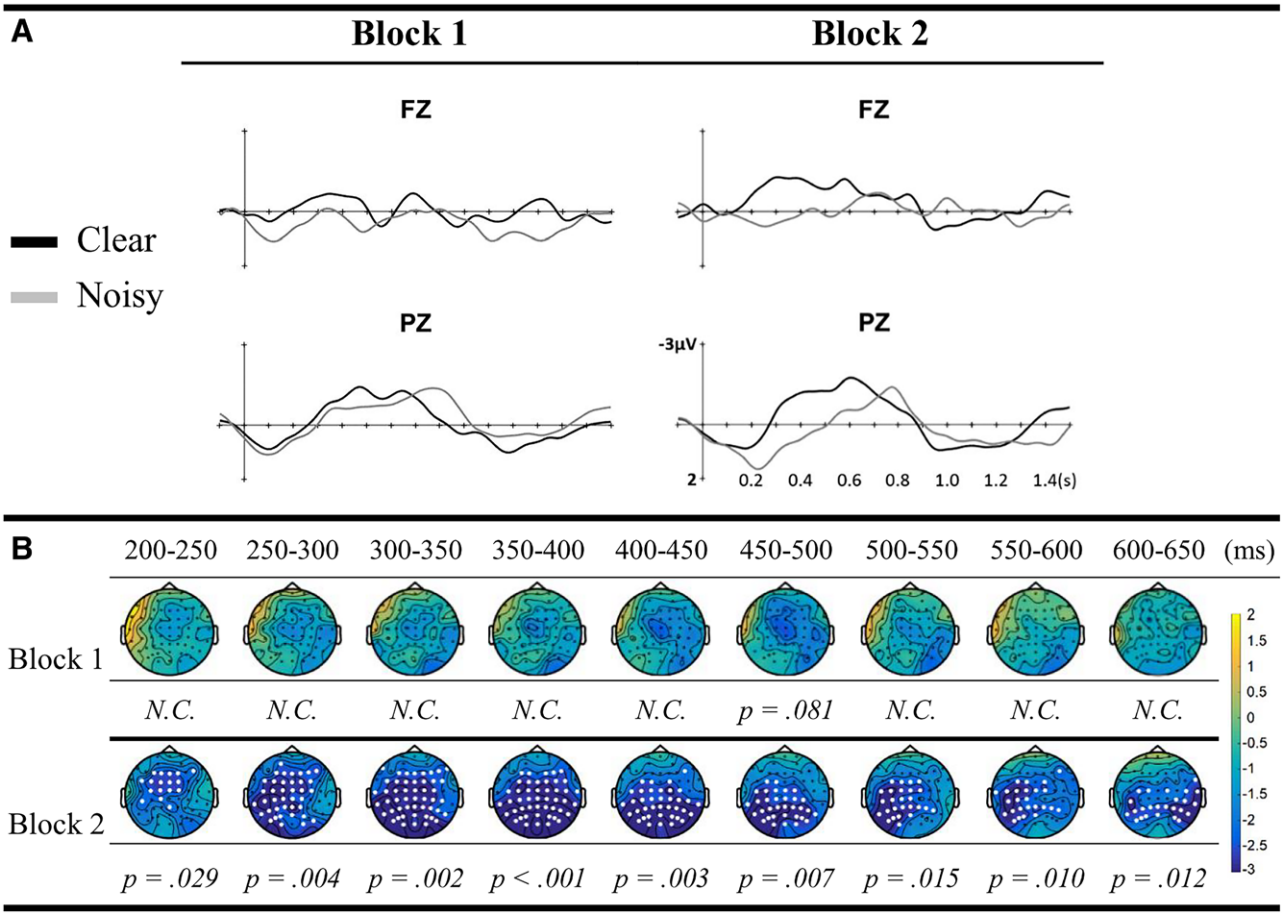

Fig. 6. Predictability effect on the N400 across clear and noisy conditions in different experimental blocks, illustrated through waveform differences and topographic distributions. A, Difference waves of the clear (black) and noisy (gray) conditions in block 1 (left) and block 2 (right). B, Topographic maps for the predictability effect in blocks 1 and 2 based on difference waves (clear—noisy) in a moving time window analysis. White dots represent electrodes with significant differences in the contrast. N.C. indicates no clusters.

Discussion

Experiment 1 investigated whether and how the N400 predictability effect manifests temporal changes that might reflect listening-related fatigue across blocks during clear and noisy sentence comprehension. The behavioral results showed ceiling comprehension accuracy in both clarity conditions, indicating that the participants maintained attention and comprehension throughout the experiment. Regarding the N400 effect, the noisy condition exhibited a reduced, delayed, and less anteriorly distributed effect compared with the clear condition (Fig. 4). When examining changes across blocks (Figs. 2, 5), we observed distinct patterns: in the clear condition, block 2 exhibited an N400 effect with an earlier onset, later offset, and more anteriorly

TABLE 1. LMM estimates of fixed effects of predictability, clarity, and block on the N400 in experiment 1

	N400 (300–850 msec)							
	Anterior				Posterior			
Variables	β	SE	*t* Value	*p*	β	SE	*t* Value	*p*
(Intercept)	−0.1580	0.1106	−1.429	0.155	−0.4299	0.0919	−4.677	<0.001
Pred.	−0.3892	0.1691	−2.302	0.023	−0.6906	0.1379	−5.007	<0.001
Clarity	0.6404	0.1461	4.383	<0.001	0.3050	0.1251	2.438	0.017
Block	−0.0912	0.0364	−2.506	0.012	−0.0304	0.0336	−0.905	0.366
Pred. × Clarity	0.5544	0.0642	8.642	<0.001	0.3945	0.0593	6.659	<0.001
Pred. × Block	−0.2653	0.0732	−3.625	<0.001	−0.0818	0.0675	−1.212	0.225
Clarity × Block	−0.4746	0.0643	−7.378	<0.001	−0.1420	0.0594	−2.390	0.017
Pred. × Clarity × Block	0.4537	0.1293	3.510	<0.001	0.5486	0.1194	4.596	<0.001

Block, the contrast between block 1 and block 2; Clarity × Block, the interaction effect between clarity and block; Clarity, the contrast between clear and noisy conditions; LMM, Linear mixed model; Pred. × Block, the interaction effect between predictability and block; Pred. × Clarity × Block, the interaction effect between predictability, clarity, and block; Pred. × Clarity, the interaction effect between predictability and clarity; Pred., the contrast between low- and high-predictability sentences.

distributed topography without amplitude reductions relative to block 1. In the noisy condition, block 2 demonstrated a more consistent N400 effect across time windows, though with a later onset than block 1. Furthermore, speech clarity differentially modulated these temporal changes (Fig. 6). While block 1 showed no significant differences between clarity conditions, the noisy condition in block 2 exhibited a reduced, delayed, and posteriorly distributed effect compared with the clear condition. These patterns suggest differences in resource allocation between clear and noisy speech processing over time, which we examine in detail later.

Clear Speech and the N400 Predictability Effect • Regarding the overall clarity effect, experiment 1 replicated the results of Hsin et al. (2023), supporting previous N400 evidence of compromised lexical predictions during degraded speech comprehension (Obleser & Kotz 2011; Strauß et al. 2013; Silcox & Payne 2021). The temporal changes in the N400 effect for clear speech showed interesting patterns. While amplitude and latency in centroparietal regions remained stable across blocks, the effect exhibited a broader scalp distribution extending more anteriorly in block 2. Previous studies have associated such anteriorly distributed N400 effects with the recruitment of additional or distinct processing resources during comprehension. In Li and Ren (2012), greater accentuation of target words elicited an N400 congruency effect with a more anterior distribution and earlier onset than normally accented words, suggesting increased allocation of attentional resources enhanced early detection of semantic incongruence. Similarly, Boudewyn et al. (2013) observed that listeners with higher WM capacity exhibited a more anteriorly distributed N400 effect, specifically over right-frontal regions, during discourse comprehension. This pattern diverged from the classic posterior distribution seen in listeners with lower WM capacity and was predicted by listening span. These findings suggest that additional cognitive resources may support the integration of contextual information during comprehension.

Although the topographic distribution of ERP effects does not directly reveal underlying neural sources, variations in distribution may reflect differences in how neural systems are engaged during processing (Boudewyn et al. 2013). Studies have demonstrated that semantic processing involves a network across anterior and posterior regions, including the prefrontal cortex, IFG, and superior/middle temporal cortices (Halgren et al. 2002; Tse et al. 2007; Federmeier & Laszlo 2009; Federmeier 2021). A functional neuroanatomical model for semantic processing suggests the N400 effect manifests from interactions where the left middle temporal gyrus supports lexical storage and access, the anterior temporal cortex and angular gyrus facilitate integration with context, and the left IFG mediates controlled retrieval of lexical representations (Lau et al. 2008). Notably, neuroimaging studies indicate the frontal cortex, particularly the left IFG, also plays a role in managing attentional and verbal WM demands (D'Arcy et al. 2004; Jacquemot & Scott 2006; Wild et al. 2012).

In our clear speech condition, the broader distribution of the N400 effect in block 2 may reflect the increased engagement of these neural systems over time. As participants continuously attended to spoken sentences, additional resources may have been recruited to maintain online processing efficiency. The earlier onset of this more broadly distributed N400 effect, without amplitude reductions, suggests this additional recruitment may help sustain efficient processing. These patterns align with findings from Li and Ren (2012), where enhanced attentional allocation was associated with earlier and more anterior N400 effects.

Evidence from nonlinguistic paradigms provides complementary insights into how neural activity patterns change during sustained task performance. Using a combined visual-auditory recognition task over 2.5 hr, Sabeti et al. (2018) found that P3b sources shifted from broad temporo-parieto-occipital, inferior frontal, and cingulate regions to more focal temporal and inferior frontal regions, while maintaining stable amplitudes. In a 160-min Stroop task, Wang et al. (2016) observed a late frontal ERP component that followed an inverted *U*-shaped trajectory, peaking between 60 and 100 min before declining. This frontal activity emerged when participants' behavioral performance worsened, but the performance was maintained during its peak amplitude period. However, the ERP amplitude and behavioral performance deteriorated in later stages. The authors interpreted this pattern as reflecting compensatory mechanisms: frontal regions were initially recruited to maintain task performance as fatigue developed, but these compensatory processes eventually failed as fatigue worsened. The study further suggested possible relationships between this frontal activity and anterior cingulate or orbitofrontal cortices that regulate cognitive control and fatigue-related compensation. Using positron emission tomography imaging during a 35-min trail-making task, Tajima et al. (2010) found increased medial orbitofrontal activation that correlated with subjective fatigue ratings despite stable performance in reaction time. The authors interpreted this increased orbitofrontal activity as reflecting a feedback mechanism that monitors and regulates effort expenditure to prevent exhaustion during prolonged task performance, suggesting this region's role in both fatigue perception and effort regulation.

Alternative explanations for the observed block effects, such as stimulus adaptation, seem less likely given the existing literature. Studies examining N400 adaptation have shown mixed results, with several finding the N400 effect remains stable despite repeated exposure to semantic anomalies (Yano et al. 2021) or changes in predictive context (Zhang et al. 2019; van Wonderen & Nieuwland 2023). When adaptation effects have been observed, as in Brothers et al. (2019), where speaker reliability was manipulated by varying the proportion of predictable versus unpredictable sentence endings between two speakers (i.e., one speaker produced mostly predictable endings while the other mostly unpredictable), the effect manifested as amplitude changes in centroparietal regions rather than the broader distribution we observed. Moreover, our experimental design, with equal numbers of high- and low-predictability sentences in randomized order, provided a limited basis for systematic adaptation.

Noisy Speech and the N400 Predictability Effect • In block 1, the noisy speech elicited an inconsistent N400 effect across time windows with a delayed onset (Fig. 5) compared with clear speech. Reductions and delays in N400 effects under noisy conditions suggest listeners' reduced ability to use context to anticipate word meaning or phonological features, relying instead on a bottom-up "wait and see" approach that hinders efficient semantic retrieval (Silcox & Payne 2021). Because cognitive capacity is limited and shared between sensory and semantic processing, processing degraded speech signals might have captured resources that would otherwise be utilized for efficient semantic processing (Aydelott et al. 2006; Strauß et al. 2013).

In block 2, the N400 effect for noisy speech showed further reductions and delays in onset latency (Fig. 6), suggesting a progressive decline in online semantic processing efficiency. In addition, noisy speech elicited primarily centroparietal N400 effects across blocks, contrasting with the more anteriorly extended effects for clear speech in block 2. These topographic patterns align with previous findings on speech comprehension under clear versus adverse conditions (Strauß et al. 2013; Grey & van Hell 2017; Hsin et al. 2023; Wambacq et al. 2023). While the N400 typically extends more anteriorly in auditory than visual modalities (Kutas & Van Petten 1994; Van Petten & Luka 2006), the functional significance of a posteriorly distributed N400 effect under adverse listening conditions remains to be fully understood. Functional MRI studies provide some insight into neural responses across different listening conditions. Left IFG activity increases with improved speech clarity, particularly for less predictable sentences, suggesting its role in supporting semantic integration as intelligibility increases (Obleser & Kotz 2010). This increased IFG activity under clearer conditions may reflect a greater allocation of neural resources toward predictive and integrative processing. In contrast, Obleser and Kotz (2010) observed activities restricted in the superior temporal sulcus and superior temporal gyrus during degraded speech processing, suggesting engagement of these specific regions for processing challenging auditory input.

An unexpected finding was that while noisy speech in block 1 exhibited an inconsistent N400 effect across time windows, suggesting disrupted semantic processing, the effect became more consistent in block 2, albeit with a further delayed onset. This pattern might reflect a learning effect occurring alongside the observed processing changes (Mattys et al. 2012). This learning effect was unexpected because we intentionally used different babble noise snippets for each sentence to avoid adaptation to repeated noise patterns. According to Mattys et al. (2012), perceptual learning—involving long-lasting adjustments in the perceptual system to enhance responses to environmental challenges—frequently occurs under adverse listening conditions. Such learning recalibrates perception based on prior experience to improve future comprehension and is most effective when supported by lexical cues or feedback. However, its effectiveness typically diminishes when adverse conditions are unpredictable or lack consistent patterns.

The more stable manifestation of the N400 effect across time windows in block 2 suggests participants may have adapted to processing speech in noise, despite varying noise snippets across trials. However, the further delayed onset in block 2 indicates slower semantic processing, possibly reflecting the increased resource demands of processing degraded speech over time. These findings suggest that excessive noise may progressively deplete resources that could otherwise support efficient online semantic processing, resulting in the observed changes in N400 effects.

The temporal changes we observed in the N400 effect align with patterns that might be expected during the development of listening-related fatigue. For clear speech, the broader distribution extending anteriorly over time, coupled with maintained amplitude and even enhanced temporal characteristics (earlier onset, later offset), parallels patterns seen in studies of cognitive fatigue where additional or distinct resources are recruited to maintain performance (Wang et al. 2016; Sabeti et al. 2018). For noisy speech, the progressive reduction and delay in the N400 effect suggest a depletion of resources that might reflect accumulated listening fatigue, similar to patterns observed when compensatory mechanisms fail under sustained cognitive load. While our experimental session was relatively brief compared to some fatigue studies, the continuous nature of our listening task, with minimal task demands beyond comprehension itself, allows us to observe the effects of sustained auditory processing more directly. The distinct patterns between clear and noisy conditions—sustained efficiency through resource recruitment versus progressive deterioration—suggest that speech clarity modulates how listeners manage resources during extended listening periods, potentially reflecting different trajectories of listening-related fatigue. Overall, experiment 1 demonstrated that the N400 predictability effect's amplitude, latency, and scalp distribution may provide insight into how resource availability affects semantic processing efficiency under different listening conditions over time. These findings suggest the potential utility of the N400 effect as an objective measure of online semantic processing changes during extended periods of speech comprehension.

EXPERIMENT 2: EVALUATING THE EFFECTIVENESS OF ENHANCED SPEECH ON LISTENING FATIGUE

Experiment 1 revealed distinct patterns in how the N400 predictability effect changes over time during speech comprehension. Clear speech maintained processing efficiency through broader resource recruitment, while noisy speech showed progressive deterioration in processing. These patterns suggest that the N400 effect's temporal characteristics of amplitude, latency, and topographic distribution might serve as objective markers for assessing sustained speech processing under different listening conditions. Building on these findings, experiment 2 aimed to evaluate whether these N400 markers could help assess the effectiveness of speech enhancement systems in supporting sustained comprehension. We hypothesized that both Transformer- and MMSE-based speech enhancement methods would improve speech processing compared with noisy speech, potentially showing N400 effect patterns more similar to those observed with clear speech. These improvements might manifest as maintained amplitudes, earlier onsets, and broader topographic distributions that could reflect more efficient resource allocation. Given their different technical approaches, we expected the state-of-the-art Transformer-based enhancement to show greater benefits than the MMSE-based enhancement in supporting sustained speech processing.

Materials and Methods

Participants • Thirty-five adults participated in the Transformer condition (15 males, aged 20 to 29 years, mean = 24 ± 3), and a different group of 35 adults participated in the MMSE condition (15 males, aged 20 to 30 years, mean = 24.2 ± 2.86). All were right-handed native Mandarin Chinese speakers (mean years of education: Transformer group = 16.11 ± 1.57; MMSE group = 16.09 ± 1.27), with normal or corrected-to-normal vision and no reported neurological/psychiatric disorders or brain damage. PTA confirmed normal-hearing thresholds below 25 dB averaged across frequencies in each participant's better ear, except for two MMSE participants with mild 32.5- and 37.5-dB hearing loss. These participants were included in the analysis as their thresholds were well below the 65-dB sound

level required in this study. No significant PTA differences were found between groups ($t = -1.22$, $p = 0.22$). All recruitment procedures complied with the ethical human research conduct regulated by the local IRB.

Stimuli and Procedure • Two enhanced conditions were created using Transformer- and MMSE-based speech enhancement models to denoise the noise-corrupted sentences from experiment 1 (Ephraim & Malah 1984; Lu et al. 2023). These models were selected to generate experimental conditions based on automated metrics commonly used to assess the performance of speech enhancement techniques, such as the Perceptual Evaluation of Speech Quality (Rix et al. 2001) and Short-time Objective Intelligibility (Taal et al. 2011). As in experiment 1, a between-subjects design was used, with each participant experiencing only one listening condition (either Transformer or MMSE enhancement). There was no overlap in participants between experiments 1 and 2. The experimental procedures, including trial structure, timing, and task requirements, were identical to those used in experiment 1. Two enhanced conditions were designed to explore whether the N400 predictability effect reflected differences in model performance for human listeners (Fig. 7).

EEG/ERP Acquisition, Preprocessing, and EEMD Analysis • The apparatus, parameters, and processing protocols used for data acquisition, preprocessing, and EEMD analysis were identical to those used in experiment 1. The trial rejection rates for the Transformer and MMSE conditions were 3.21 and 3.15%, respectively. Hilbert spectrum analysis revealed that the central frequencies of IMF5 to IMF8 were 11, 5, 2, and 1 Hz.

Statistical Analysis • Figure 8 presents the grand-averaged ERMs elicited by low- and high-predictability words, with the N400 component visible from 400 to 900 msec in the centroparietal regions. Experiment 2 aimed to determine whether enhanced speech elicits N400 effects similar to those observed with clear speech, as evidence of mitigated listening-related fatigue. The data were split into blocks corresponding to the first and second halves of the experiment. To ensure consistent data quality across blocks, we examined trial retention: both MMSE and Transformer conditions showed consistent retention with an average of 62 trials in both blocks. CBPTs were conducted to characterize the temporal dynamics and topographic distributions of the predictability effects. Subsequently, the LMM analysis examined whether enhanced speech modulated the N400 effect on listening-related fatigue.

Results

Behavioral Performance • Participants in the Transformer and MMSE conditions achieved overall mean accuracies of 97.37% (SD = 3.74, range = 83.34 to 100%) and 96.98% (SD = 5.84, range = 75 to 100%), on the comprehension questions respectively. A two-way repeated-measures analysis of variance was conducted to analyze the effects of predictability and clarity on accuracy. The results showed no significant interaction between predictability and clarity ($F_{1, 68} = 1.00$, $p = 0.321$). However, there was a significant main effect of predictability ($F_{1, 68} = 7.23$, $p = 0.009$), with high-predictability sentences (mean = 97.9%, SD = 4.47) showing higher accuracy than low-predictability sentences (mean = 96.5%, SD = 5.21) across enhancement conditions. There was no significant main effect of clarity ($F_{1, 68} = 0.14$, $p = 0.711$).

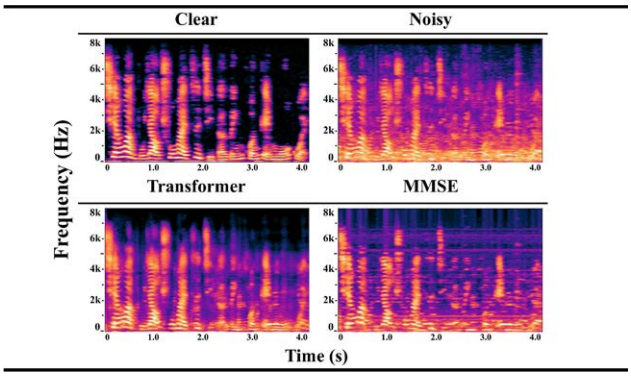

Fig. 7. Spectrograms of the example sentence—*Never sell your soul to the devil*—in the clear (upper left), noisy (upper right), Transformer-enhanced (lower left), and MMSE-enhanced (lower right) conditions. MMSE indicates minimum mean square error.

Cluster-Based Permutation Tests • The predictability effect (difference between low- and high-predictability conditions) varied between enhancement conditions across blocks (Fig. 9). In the Transformer condition, block 1 showed significant negative clusters in centroparietal regions from 500 to 750 msec ($ps < 0.05$), while block 2 showed significant negative clusters in centroparietal and frontocentral regions from 700 to 850 msec ($ps < 0.05$).

In the MMSE condition, block 1 showed significant negative clusters in left-parietal regions from 500 to 650 msec ($ps < 0.05$) and in parieto-occipital regions from 800 to 850 msec ($p = 0.048$). Block 2 showed significant negative clusters in frontal and centroparietal regions from 600 to 800 msec ($ps < 0.05$). Compared with the Transformer condition, the MMSE condition elicited a less consistent N400 effect across time windows, with a later offset in block 1. Both enhancement conditions exhibited a delayed onset of the N400 effect in block 2 compared with block 1.

To test the interaction between enhancement conditions across blocks, we analyzed difference waves, which revealed no significant clusters in either block ($ps > 0.05$). While block 2 results showed limited benefits of enhanced speech in alleviating listening fatigue, the Transformer-enhanced speech in block 1 demonstrated a temporally consistent N400 predictability effect in the centroparietal regions from 500 to 750 msec, which might suggest improved clarity for semantic processing. We therefore conducted an LMM analysis to compare the predictability effects between enhanced and noisy speech in block 1.

LMM Analysis • Table 2 summarizes the LMM analysis examining predictability and clarity effects on N400s in the posterior ROI during block 1.

Transformer Versus Noisy Conditions •

Two-way interaction Predictability × Clarity. The analysis revealed a significant interaction between predictability and clarity ($\beta = 0.3991$, SE = 0.0946, $p < 0.001$). Post hoc analysis showed that the predictability effect was significant in the Transformer condition ($\beta = -0.909$, SE = 0.232, z ratio = −3.911, $p < 0.001$), with low-predictability sentences eliciting greater

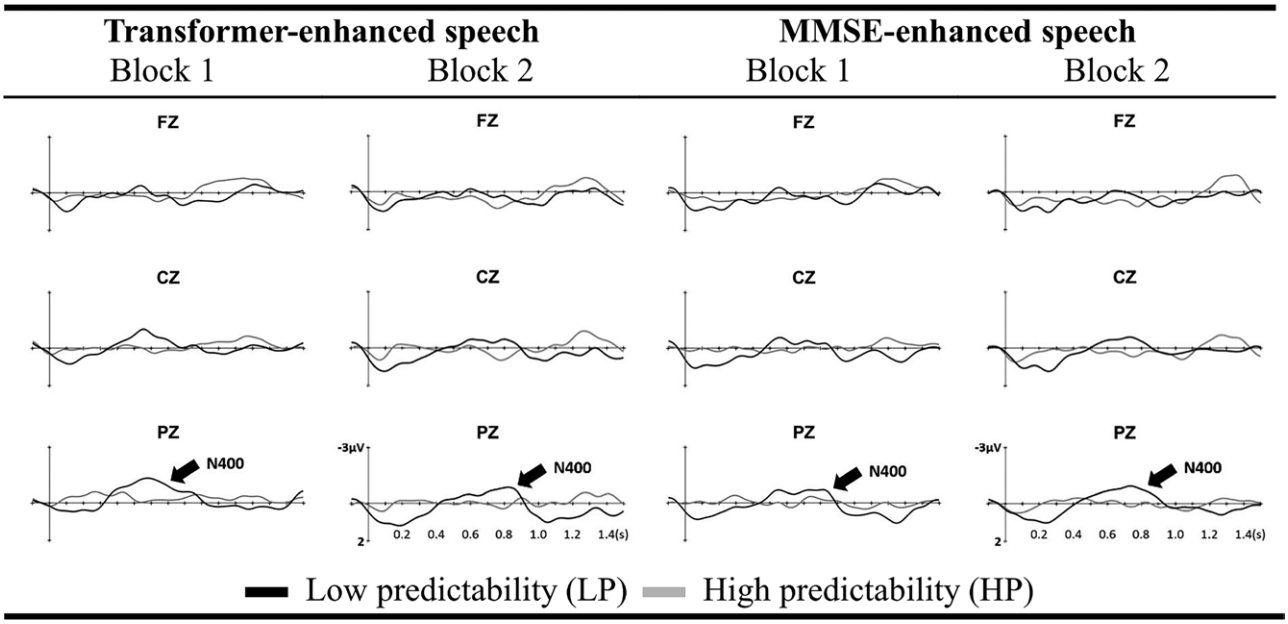

Fig. 8. Averaged ERMs of the low- (black) and high- (gray) predictability words in blocks 1 and 2 under the Transformer (left) and MMSE (right) conditions. ERMs indicate event-related modes.

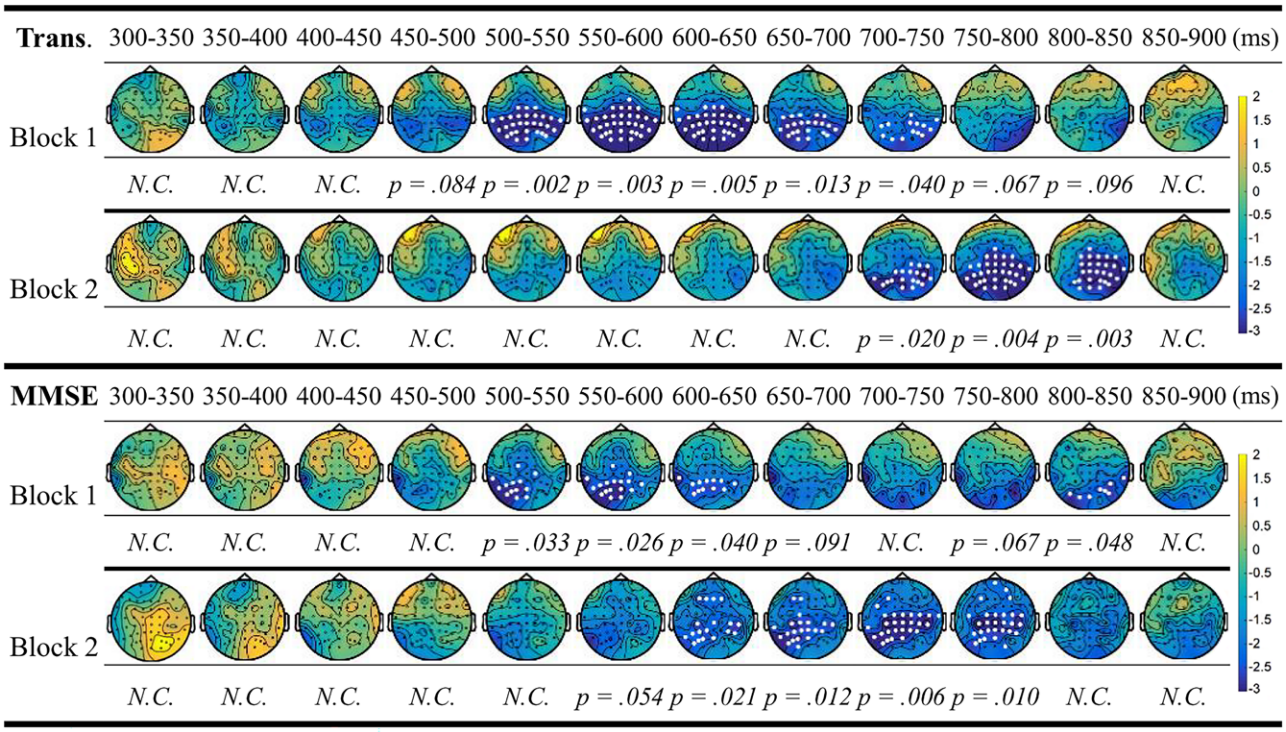

Fig. 9. Topographic maps for the predictability effect on the N400 in blocks 1 and 2 under the Transformer (upper panel) and MMSE (lower panel) conditions in a moving time window analysis. White dots represent electrodes with significant differences in the contrasts. MMSE indicates minimum mean square error; N.C., no clusters; Trans, Transformer.

N400 negativity than high-predictability sentences. This effect was not significant in the noisy condition ($p = 0.079$).

Main effects. A significant main effect of predictability emerged ($\beta = -0.7091$, SE = 0.2273, $p = 0.002$), with low-predictability sentences eliciting greater N400 negativity than high-predictability sentences. The main effect of clarity was not significant ($p = 0.731$).

MMSE versus noisy conditions •

Two-way interaction Predictability × Clarity. No significant interaction emerged between predictability and clarity ($p = 0.767$).

Main effects. The analysis showed a significant main effect of predictability ($\beta = -0.6350$, SE = 0.1933, $p = 0.001$), with low-predictability sentences eliciting greater N400 negativity

TABLE 2. LMM estimates of fixed effects of predictability and clarity on the N400 in the posterior ROI under conditions of noisy and enhanced speech for block 1

	N400 (500–750 msec)							
	Transformer Vs. Noisy				MMSE Vs. Noisy			
Variables	β	SE	*t* Value	*p*	β	SE	*t* Value	*p*
(Intercept)	−0.3680	0.1520	−2.422	0.017	−0.3257	0.1379	−2.361	0.019
Pred.	−0.7091	0.2273	−3.120	0.002	−0.6350	0.1933	−3.284	0.001
Clarity	0.7141	0.2070	0.345	0.731	−0.0804	0.2018	−0.398	0.692
Pred. × Clarity	0.3991	0.0946	4.219	<0.001	−0.0272	0.0918	−0.296	0.767

Clarity, the contrast between Transformer/MMSE and noisy conditions; LMM, Linear mixed model; MMSE, minimum mean square error; Pred. × Clarity, the interaction effect between predictability and clarity; Pred., the contrast between low- and high-predictability sentences; ROI, regions of interest.

than high-predictability sentences. The main effect of clarity was not significant ($p = 0.692$).

Discussion

Experiment 2 investigated whether the N400 predictability effect could serve as an objective measure to assess the benefits of speech enhancement systems during comprehension of Transformer- and MMSE-enhanced speech. The behavioral results showed nearly ceiling comprehension accuracy in both speech enhancement conditions, indicating well-maintained attention and comprehension throughout the experiment. We observed that both enhanced conditions showed delayed N400 effects in centroparietal regions in block 2 compared with block 1 (Fig. 9), but importantly, Transformer-enhanced speech in block 1 elicited a significantly increased N400 effect compared with noisy speech (Table 2). These findings suggest that while speech enhancement techniques may offer immediate benefits for speech processing, these advantages may not be sustained over time.

The temporal patterns of the N400 effect provide insights into how different speech enhancement approaches affect speech processing. For Transformer-enhanced speech, block 1 results demonstrated that effective enhancement could reduce disruptions to semantic processing, likely by decreasing the cognitive demands of auditory processing. This allowed more resources to be allocated toward semantic processing, as evidenced by the increased N400 effect. However, model limitations may affect long-term processing efficiency. Current speech enhancement techniques can introduce artifacts such as distortion and residual noise (Das et al. 2020), making noise patterns less predictable for listeners (Mattys et al. 2012). While listeners can mentally suppress residual noise and restore phonemes (Sivonen et al. 2006; Strauß et al. 2014), these additional cognitive operations may contribute to increased processing demands over time.

For MMSE-enhanced speech, block 1 results revealed an inconsistent N400 effect across time windows, suggesting disrupted semantic processing possibly due to less effective noise reduction. Similar to Transformer-enhanced speech, the MMSE condition showed delayed N400 effect onset latency in block 2. However, this delay was more moderate than in the Transformer condition, which might reflect the preserved consistency of noise patterns allowing for some adaptation to the listening conditions.

These findings extend previous ERP research on speech enhancement effectiveness in several important ways. MMN studies have shown that noise suppression algorithms can improve early auditory processing (Yu et al. 2018, 2019). While these MMN findings demonstrated speech enhancement benefits at the preattentive level of acoustic change detection, our N400 results reveal speech enhancement effects on higher-order semantic processing. The Transformer-enhanced speech initially facilitated semantic processing, as shown by the increased N400 effect in block 1, complementing previous observations of improved early auditory processing. However, this benefit was not observed with MMSE-enhanced speech, which showed disrupted semantic processing even in block 1.

The block analysis revealed the limitations of current speech enhancement approaches. Neither enhancement method maintained its processing advantages over time, as evidenced by delayed N400 effects in block 2, similar to the patterns we observed with noisy speech in experiment 1. The fact that even the better-performing Transformer enhancement showed these temporal changes suggests current speech enhancement technology may face inherent limitations in supporting sustained listening. These limitations might stem from the additional cognitive operations needed to process enhanced but imperfect signals, as listeners must still engage in noise suppression and phoneme restoration despite the enhancement. While speech enhancement systems may effectively improve immediate speech processing, as shown by the initial benefits of Transformer enhancement, the progressive delays in N400 effects suggest these benefits may not protect against the resource demands of extended listening. These findings suggest that future development of speech enhancement systems might benefit from considering not only immediate processing benefits but also how enhancement strategies affect sustained listening over time. Overall, experiment 2 demonstrated that the N400 predictability effect may serve as a sensitive measure for evaluating both the immediate benefits of speech enhancement systems and their limitations in supporting sustained speech processing. The distinct patterns between enhancement methods, particularly in block 1, suggest this measure can also capture meaningful differences in speech enhancement model performance.

GENERAL DISCUSSION

The present study examined how the N400 predictability effect changes over time during speech comprehension under clear, noisy, and enhanced conditions. Our aims were twofold: to investigate whether temporal changes in the N400 effect's amplitude, latency, and topography could serve as objective markers of listening-related fatigue, and to evaluate whether these markers could assess the effectiveness of speech enhancement systems. By measuring N400s elicited by sentence-final words, we found distinct patterns that suggest how listening-related fatigue may affect resource allocation and semantic processing under different listening conditions. These findings point to potential

applications of the N400 effect for quantifying the cognitive impact of listening-related fatigue in real-world scenarios.

Listening-Related Fatigue and Resource Allocation in Speech Comprehension

Recent research increasingly recognizes the importance of considering both peripheral and cognitive factors in hearing assessments, particularly in noisy environments (Committee on Hearing, Bioacoustics, & Biomechanics [CHABA] 1988; Arlinger et al. 2009; Pichora-Fuller et al. 2016; Peelle 2018; Ronnberg et al. 2019). While PTA remains a standard assessment tool, it may not accurately reflect listeners' speech understanding capabilities or their susceptibility to listening-related fatigue (Hind et al. 2011; Ruggles et al. 2011). Studies have demonstrated that the ability to utilize contextual information plays a crucial role in comprehending speech under adverse conditions (Kalikow et al. 1977; Mattys et al. 2012; Obleser 2014; Payne & Silcox 2019). ERP research has established that this context effect reflects an active prediction of upcoming lexical information to facilitate semantic processing (Kutas & Federmeier 2011). However, few EEG/ERP studies have used this predictive processing framework to investigate how listening-related fatigue and effort affect semantic processing during sentence comprehension (Obleser & Kotz 2011; Song & Iverson 2018; Hunter 2020, 2022), leaving questions about the neurocognitive mechanisms through which fatigue impacts speech comprehension.

On the basis of lexical predictions, the present study shows that the N400 effect manifests distinct temporal patterns under clear and noisy conditions, suggesting its potential utility as a marker of listening-related fatigue. When processing speech in excessive noise, we observed both decreased amplitude differences between predictable and unpredictable words and delayed onset of these differences. In contrast, clear speech maintained efficient predictive processing through the recruitment of additional resources, as indicated by broader topographic distributions. These differences emerged despite our participants being young adults with typical hearing acuity, highlighting how listening conditions alone can affect resource allocation patterns.

The observed patterns likely reflect how different listening demands affect cognitive resource utilization. According to limited-capacity models, cognitive resources are shared across processing domains (Kahneman 1973; Strauß et al. 2013; Pichora-Fuller et al. 2016). With its lower processing demands, clear speech may allow listeners to recruit additional resources to maintain efficiency as fatigue develops. Conversely, the high demands of processing noisy speech may quickly approach capacity limits, leaving few resources available for semantic processing. This resource limitation becomes evident in the reduced ability to utilize contextual information and the progressive deterioration of processing efficiency over time.

Notably, attentive listening to clear speech for approximately 30 min elicited topographic changes in the N400 effect, suggesting altered patterns of resource allocation. While we interpret these changes cautiously, previous studies have demonstrated that mental fatigue can emerge during both active and passive tasks lasting 30 to 35 min. Tajima et al. (2010) found increased fatigue ratings and medial orbitofrontal activity after a 35-min trail-making test, even when behavioral performance remained stable. Similarly, Lorcery et al. (2024) examined mental fatigue effects on effort engagement by comparing psychophysiological and behavioral responses during a 30-min Stroop task with a non-fatiguing control session of documentary viewing. During the Stroop task, participants showed clear signs of mental fatigue, including increased reaction times and higher omission rates that aligned with reduced sympathetic activity, indicated by an increased pre-ejection period, a cardiographic measure signaling effort disengagement associated with mental fatigue. Notably, the control condition showed a smaller but significant increase in the pre-ejection period without corresponding behavioral changes, suggesting that minimally demanding tasks can elicit mild fatigue responses. These studies support the possibility that our observed N400 changes may reflect fatigue-related alterations in resource allocation, even during relatively brief listening periods.

Perceptual Learning and Adaptation During Noisy Speech Processing

Listeners in the noisy condition exhibited an interesting pattern. While their semantic processing efficiency decreased, they showed evidence of noise adaptation through perceptual learning. Mattys et al. (2012) propose that perceptual learning frequently occurs under adverse listening conditions. When participants are required to listen attentively and provide behavioral responses, adaptation through learning may occur more frequently for goal-oriented comprehension within the experimental context (Francis & Love 2019; Federmeier 2021). Previous ERP studies have observed such adaptation during foreign-accented speech comprehension, where listeners showed improved semantic processing and increased tolerance to syntactic errors over time (Hanulíková et al. 2012; Romero-Rivas et al. 2015). In this study, although we used different noise snippets for each sentence to minimize learning effects, the consistent noise type may have allowed listeners to develop adaptive strategies such as postdiction (Ronnberg et al. 2019; Winn & Teece 2021), semantic matching (Chwilla et al. 1998; Lau et al. 2013), or inference-based correction (Ryskin et al. 2021; Hsin et al. 2023). However, this adaptation process may come with a cost. The more consistent but increasingly delayed N400 effects suggest that while listeners can adapt to noise, this adaptation may contribute to increased processing demands and fatigue over time.

Evaluating Speech Enhancement: From Current Methods to Future Directions

Current methods for evaluating speech enhancement systems rely primarily on automated metrics and subjective ratings, which often show limited correlation (Zezario et al. 2023). While automated metrics may not fully capture user experience, behavioral measures, including subjective ratings, reflect the cumulative effects of various cognitive processes and task-related decision-making (Lau et al. 2013). ERPs offer a valuable alternative by providing high temporal resolution measurements of ongoing cognitive processing, allowing the detection of subtle differences in how various speech enhancement systems affect speech processing over time.

During the first half of the experiment, Transformer-enhanced speech elicited a significantly increased N400 effect relative to noisy speech, indicating the short-term effectiveness of the

speech enhancement model in improving semantic processing. However, the artifacts inherent in current speech enhancement techniques may limit their long-term benefits in alleviating listening-related fatigue. These findings align with Ohlenforst et al. (2017), who demonstrated the limited benefits of hearing aids in reducing listening effort. Consequently, users of assistive hearing devices may be less motivated to develop healthy usage habits when long-term fatigue outweighs short-term processing benefits (Eckert et al. 2016), highlighting the importance of considering user experience over time when evaluating system effectiveness.

The second half of the experiment revealed another important pattern. MMSE-enhanced speech appeared to elicit some noise adaptation compared with Transformer-enhanced speech. This observation highlights core challenges in developing effective speech enhancement systems. Higher-performing models like the Transformer achieve better noise reduction but risk introducing distortions, while suboptimal models maintain greater residual noise but more consistent patterns. The former provides immediate improvements in speech comprehension but may accelerate fatigue development through the processing demands of managing distortions. The latter allows for adaptation but at the cost of continued exposure to residual noise. It is interesting that listeners adapted more effectively to unenhanced noisy speech than enhanced speech, suggesting that current enhancement methods might interfere with natural adaptation processes. This finding could inform future refinements of speech enhancement systems to better balance immediate processing benefits with support for sustained listening.

These findings have important implications for developing and evaluating hearing assistance technologies. While current speech enhancement systems can effectively improve immediate speech processing, their benefits may not extend to alleviating the cognitive demands of sustained listening. Future technological developments might benefit from considering not only immediate intelligibility improvements but also how enhancement strategies affect listening-related fatigue over extended periods. The N400 effect appears to provide a sensitive measure for assessing both these immediate and longer-term aspects of speech enhancement system performance.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations of the present study should be noted. First, this study represents an exploratory investigation of using the N400 and sentence stimuli to examine listening-related fatigue. Although our interpretation is supported by both our ERP findings and previous literature, the lack of direct behavioral measures of fatigue suggests the need for additional validation. However, future studies should carefully consider the selection of behavioral measures, as research has shown limited correlations between different types of fatigue/effort measurements (McGarrigle et al. 2017; Decruy et al. 2020; Kestens et al. 2023; for a review, see Shields et al. 2023). Second, we examined only one type of noise at a single SNR level. Further research incorporating various noise types and SNR levels is needed to fully evaluate the viability of the N400 as a fatigue measurement tool. Third, our sample was limited to young adults with normal hearing, and the results may not generalize to older adults or individuals with hearing impairment who are more likely to experience listening fatigue in daily life.

Methodologically, while CBPT effectively controls type I error for detecting differences between conditions, their limitations in establishing precise temporal and spatial parameters should be noted (Sassenhagen & Draschkow 2019). These tests provide weak control over type I error rate at individual time points or locations (Luck 2014). Therefore, while our cluster-based analyses reliably indicate the presence of condition differences, the temporal windows and topographical distributions we report should be interpreted as descriptive characteristics of the observed data rather than statistically precise estimates of effect timing and location. This statistical consideration aligns with our approach of reporting approximate time windows and broad topographical distributions rather than specific onset/offset times or precise spatial locations.

Regarding sample characteristics, the uneven gender ratio in our participant groups, with more females than males, highlights the importance of considering gender differences in future research. These differences have been documented in auditory processing and hearing loss risk (Kraus 2019; Reavis et al. 2023). Future research could examine whether these biological differences contribute to listening-related fatigue, informing gender-specific auditory assessments and interventions.

CONCLUSIONS

The present study offers novel insights into how listening-related fatigue affects higher-order language comprehension, using the N400 predictability effect as a potential neural marker. Through examining speech comprehension under clear, noisy, and enhanced conditions, we demonstrated that temporal changes in the N400 effect's amplitude, latency, and scalp distribution may reflect how listeners manage cognitive resources during sustained listening. In clear speech, listeners appear to maintain processing efficiency by recruiting additional resources, while excessive noise progressively compromises semantic processing despite some adaptation. When applied to enhanced speech, this measure revealed that while current speech enhancement technologies can provide immediate processing benefits, they may face limitations in supporting sustained listening. These findings contribute to cognitive hearing science by providing a potential objective measure of listening-related fatigue during real-world communication. Moreover, the N400 effect shows promise as a tool for evaluating hearing assistance technologies, potentially guiding the development of interventions that better support both immediate comprehension and sustained listening. Future applications of this measure might help identify individuals at risk for hearing-related cognitive difficulties, enabling timely intervention to maintain healthy communication practices.

ACKNOWLEDGMENTS

The authors of this study wish to express their sincere gratitude to Yu-Anne Shiah for recording the experimental stimuli and assisting with data collection, to Ching-Chih John Sung, Kuo-Hsuan Hung, and You-Jin Li for generating denoised stimuli, and to the Brain and Language Lab at Academia Sinica for their technical support. Furthermore, the authors extend their heartfelt thanks to all the study participants.

This research was supported by grants from Academia Sinica: AS-HLGC-112-02. The funding organization had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the

data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

C.-H.H conceived the study, designed the experiments, collected and analyzed the data, interpreted the results, wrote the manuscript, and acquired the research funding. C.-Y.L. designed the experiments and interpreted the results. Y.T. conceived the study, designed the experiments, proofread the manuscript, and acquired the research funding.

This submitted manuscript reports data derived from experimental observations in human subjects. This study was approved by the Academia Sinica Institutional Review Board for Biomedical Science Research (IRB-BM), with application number: AS-IRB-BM-18039 (R1). The Principal Investigator's name submitted to the IRB for review: C.-Y. L.

The authors have no conflicts of interest to disclose.

Address for correspondence: Yu Tsao, Biomedical Acoustic Signal Processing Lab, Research Center for Information Technology Innovation, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan. E-mail: yu.tsao@citi.sinica.edu.tw

Received July 23, 2024; accepted December 05, 2024; published online ahead of print February 19, 2025

REFERENCES

Arlinger, S., Lunner, T., Lyxell, B., Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scand J Psychol*, *50*, 371–384.

Aydelott, J., Dick, F., Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*, 454–464.

Baayen, R., Davidson, D., Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*, *59*, 390–412.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software*, *67(1)*, 1–48.

Bertoli, S., & Bodmer, D. (2014). Novel sounds as a psychophysiological measure of listening effort in older listeners with and without hearing loss. *Clin Neurophysiol*, *125*, 1030–1041.

Bertoli, S., & Bodmer, D. (2016). Effects of age and task difficulty on ERP responses to novel sounds presented during a speech-perception-in-noise test. *Clin Neurophysiol*, *127*, 360–368.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res*, *27*, 32–48.

Boudewyn, M. A., Long, D. L., Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Front Psychol*, *4*, 60.

Boudewyn, M. A., Long, D. L., Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cogn Affect Behav Neurosci*, *15*, 607–624.

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, 107225.

Brothers, T., Swaab, T. Y., Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *J Mem Lang*, *93*, 203–216.

Burkhardt, P., Muller, V., Meister, H., Weglage, A., Lang-Roth, R., Walger, M., Sandmann, P. (2022). Age effects on cognitive functions and speech-in-noise processing: An event-related potential study with cochlear-implant users and normal-hearing listeners. *Front Neurosci*, *16*, 1005859.

Committee on Hearing, Bioacoustics, & Biomechanics (CHABA). (1988). Speech understanding and aging. *J Acoust Soc Am*, *83*, 859–895.

Chang, C. T., Lee, C. Y., Chou, C. J., Fuh, J. L., Wu, H. C. (2016). Predictability effect on N400 reflects the severity of reading comprehension deficits in aphasia. *Neuropsychologia*, *81*, 117–128.

Chao, P.-C., Chen, W.-F., Lee, C.-Y. (2019). The second-order effect of orthography-to-phonology mapping consistency on Chinese spoken word recognition. *J Neurolinguistics*, *51*, 1–16.

Chen, W. F., Chao, P. C., Chang, Y. N., Hsu, C. H., Lee, C. Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain Lang*, *157–158*, 51–62.

Chwilla, D. J., Hagoort, P., Brown, C. M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *QJEP*, *51*, 531–560.

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., McRae, K. (2014). The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. *Neuroimage Clin*, *4*, 788–799.

D'Arcy, R. C., Ryner, L., Richter, W., Service, E., Connolly, J. F. (2004). The fan effect in fMRI: Left hemisphere specialization in verbal working memory. *Neuroreport*, *15*, 1851–1855.

Das, N., Chakraborty, S., Chaki, J., Padhy, N., Dey, N. (2020). Fundamentals, present and future perspectives of speech enhancement. *Int J Speech Technol*, *24*, 883–901.

Deal, J. A., Betz, J., Yaffe, K., Harris, T., Purchase-Helzner, E., Satterfield, S., Pratt, S., Govil, N., Simonsick, E. M., Lin, F. R.; Health ABC Study Group. (2017). Hearing impairment and incident dementia and cognitive decline in older adults: The health ABC study. *J Gerontol A Biol Sci Med Sci*, *72*, 703–709.

Decruy, L., Lesenfants, D., Vanthornhout, J., Francart, T. (2020). Top-down modulation of neural envelope tracking: The interplay with behavioral, self-report and neural measures of listening effort. *Eur J Neurosci*, *52*, 3375–3393.

Eckert, M. A., Teubner-Rhodes, S., Vaden, K. I. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear & Hearing*, *37*, 101S–110S.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans Acoust Speech Signal Process*, *32*, 1109–1121.

Erlbeck, H., Kubler, A., Kotchoubey, B., Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Front Hum Neurosci*, *8*, 654.

Federmeier, K. D. (2021). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, *59*, e13940.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychol Learn Motiv Adv Res Theory*, *51*, 1–44.

Francis, A. L., & Love, J. (2019). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdiscip Rev Cogn Sci*, *11*, e1514.

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *J Neurolinguistics*, *42*, 93–108.

Gustafson, S. J., Key, A. P., Hornsby, B. W. Y., Bess, F. H. (2018). Fatigue related to speech processing in children with hearing loss: Behavioral, subjective, and electrophysiological measures. *J Speech Lang Hear Res*, *61*, 1000–1011.

Gygi, B., & Ann Hall, D. (2016). Background sounds and hearing-aid users: A scoping review. *Int J Audiol*, *55*, 1–10.

Hahne, A., Wolf, A., Müller, J., Mürbe, D., Friederici, A. D. (2012). Sentence comprehension in proficient adult cochlear implant users: On the vulnerability of syntax. *Lang Cognit Process*, *27*, 1192–1204.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*, 1101–1116.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *J Cogn Neurosci*, *24*, 878–887.

Hind, S. E., Haines-Bazrafshan, R., Benton, C. L., Brassington, W., Towle, B., Moore, D. R. (2011). Prevalence of clinical referrals having hearing thresholds within normal limits. *Int J Audiol*, *50*, 708–716.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*, 286–300.

Holt, R., Kung, C., Demuth, K. (2018). Listener characteristics modulate the semantic processing of native vs. foreign-accented speech. *PLoS One*, *13*, e0207452.

Hong, J. Y., Oh, I. H., Jung, T. S., Kim, T. H., Kang, H. M., Yeo, S. G. (2014). Clinical reasons for returning hearing aids. *Korean J Audiol*, *18*, 8–12.

Hornsby, B. W., Naylor, G., Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear Hear*, *37*(Suppl 1), 136S–144S.

Hsin, C. H., Chao, P. C., Lee, C. Y. (2023). Speech comprehension in noisy environments: Evidence from the predictability effects on the N400 and LPC. *Front Psychol*, *14*, 1105346.

Hsu, C. H., Lee, C. Y., Liang, W. K. (2016). An improved method for measuring mismatch negativity using ensemble empirical mode decomposition. *J Neurosci Methods*, *264*, 78–85.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A*, *454*, 903–995.

Hunter, C. R. (2020). Tracking cognitive spare capacity during speech perception with EEG/ERP: Effects of cognitive load and sentence predictability. *Ear Hear*, *41*, 1144–1157.

Hunter, C. R. (2022). Listening over time: Single-trial tonic and phasic oscillatory alpha-and theta-band indicators of listening-related fatigue. *Front Neurosci*, *16*, 915349.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci*, *10*, 480–486.

Jamison, C., Aiken, S. J., Kiefte, M., Newman, A. J., Bance, M., Sculthorpe-Petley, L. (2016). Preliminary investigation of the passively evoked N400 as a tool for estimating speech-in-noise thresholds. *Am J Audiol*, *25*, 344–358.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Inc.

Kalikow, D. N., Stevens, K. N., Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*, *61*, 1337–1351.

Kemp, A., Eddins, D., Shrivastav, R., Hampton Wray, A. (2019). Effects of task difficulty on neural processes underlying semantics: An event-related potentials study. *J Speech Lang Hear Res*, *62*, 367–386.

Kestens, K., Van Yper, L., Degeest, S., Keppler, H. (2023). The P300 auditory evoked potential: A physiological measure of the engagement of cognitive systems contributing to listening effort? *Ear Hear*, *44*, 1389–1403.

Key, A. P., Gustafson, S. J., Rentmeester, L., Hornsby, B. W. Y., Bess, F. H. (2017). Speech-processing fatigue in children: Auditory event-related potential and behavioral measures. *J Speech Lang Hear Res*, *60*, 2090–2104.

Kraus, N. (2019). Sex differences in sound processing. *Hear J*, *72*, 6.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci*, *4*, 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol Psychol*, *11*, 99–116.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83–143). Academic Press.

Lai, M. K., Payne, B. R., Federmeier, K. D. (2023). Graded and ungraded expectation patterns: Prediction dynamics during active comprehension. *Psychophysiology*, *61*, e14424.

Lau, E. F., Holcomb, P. J., Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *J Cogn Neurosci*, *25*, 484–502.

Lau, E. F., Phillips, C., Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nat Rev Neurosci*, *9*, 920–933.

Ledwidge, P. S., Jones, C. M., Huston, C. A., Trenkamp, M., Bator, B., Laeng, J. (2022). Electrophysiology reveals cognitive-linguistic alterations after concussion. *Brain Lang*, *233*, 105166.

Li, X. Q., & Ren, G. Q. (2012). How and when accentuation influences temporally selective attention and subsequent semantic processing during online spoken language comprehension: An ERP study. *Neuropsychologia*, *50*, 1882–1894.

Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., Ferrucci, L. (2011). Hearing loss and incident dementia. *Arch Neurol*, *68*, 214–220.

Lin, F. R., Pike, J. R., Albert, M. S., et al. ACHIEVE Collaborative Research Group. (2023). Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): A multicentre, randomised controlled trial. *Lancet*, *402*, 786–797.

Liu, C. M., & Lee, C. T. (2019). Association of hearing loss with dementia. *JAMA Netw Open*, *2*, e198112.

Livingston, G., Sommerlad, A., Orgeta, V., et al. (2017). Dementia prevention, intervention, and care. *Lancet*, *390*, 2673–2734.

Lorcery, A., Andre, N., Benraiss, A., Pingault, M., Mirabelli, F., Audiffren, M. (2024). Engagement of mental effort in response to mental fatigue: A psychophysiological analysis. *Psychol Sport Exerc*, *74*, 102660.

Lu, Y.-J., Chang, C.-Y., Yu, C., Liu, C.-F., Hung, J.-w, Watanabe, S., Tsao, Y. (2023). Improving speech enhancement performance by leveraging contextual broad phonetic class information. *IEEE/ACM Trans Audio Speech Lang Process*, *31*, 2738–2750.

Luck, S. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*, *164*, 177–190.

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

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

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol*, *53*, 433–440.

Moore, T. M., Key, A. P., Thelen, A., Hornsby, B. W. Y. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia*, *106*, 371–382.

Obleser, J. (2014). Putting the listening brain in context. *Lang Linguist Compass*, *8*, 646–658.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex*, *20*, 633–640.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*, 713–723.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear Hear*, *38*, 267–281.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*, *2011*, 156869.

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. *Psychol Learn Motiv*, *71*, 215–264.

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

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear*, *37*, 5S–27S.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*, *97*, 593–608.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clin Neurophysiol*, *118*, 2128–2148.

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

Reavis, K. M., Bisgaard, N., Canlon, B., Dubno, J. R., Frisina, R. D., Hertzano, R., Humes, L. E., Mick, P., Phillips, N. A., Pichora-Fuller, M. K., Shuster, B., Singh, G. (2023). Sex-linked biology and gender-related research is essential to advancing hearing health. *Ear Hear*, *44*, 10–27.

Rix, A. W., Beerends, J. G., Hollier, M. P., Hekstra, A. P. (2001). *Perceptual Evaluation of Speech Quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs*. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221).

Romero-Rivas, C., Martin, C. D., Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Front Hum Neurosci*, *9*, 167.

Romero-Rivas, C., Martin, C. D., Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255.

Ronnberg, J., Holmer, E., Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int J Audiol*, *58*, 247–261.

Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc Natl Acad Sci U S A*, *108*, 15516–15521.

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, *158*, 107855.

Sabeti, M., Boostani, R., Rastgar, K. (2018). How mental fatigue affects the neural sources of P300 component. *J Integr Neurosci*, *17*, 71–81.

Sarant, J. Z., Busby, P. A., Schembri, A. J., Fowler, C., Harris, D. C. (2024). ENHANCE: A comparative prospective longitudinal study of cognitive outcomes after 3 years of hearing aid use in older adults. *Front Aging Neurosci*, *15*, 1302185.

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335.

Sheldon, S., Pichora-Fuller, M. K., Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J Acoust Soc Am*, *123*, 489–499.

Shields, C., Sladen, M., Bruce, I. A., Kluk, K., Nichani, J. (2023). Exploring the correlations between measures of listening effort in adults and children: A systematic review with narrative synthesis. *Trends Hear*, *27*, 23312165221137116.

Silcox, J. W., & Payne, B. R. (2021). The costs (and benefits) of effortful listening on context processing: A simultaneous electrophysiology, pupillometry, and behavioral study. *Cortex*, *142*, 296–316.

Sivonen, P., Maess, B., Lattner, S., Friederici, A. D. (2006). Phonemic restoration in a sentence context: Evidence from early and late ERP effects. *Brain Res*, *1121*, 177–189.

Skagerstrand, A., Stenfelt, S., Arlinger, S., Wikstrom, J. (2014). Sounds perceived as annoying by hearing-aid users in their daily soundscape. *Int J Audiol*, *53*, 259–269.

Smeds, K., Wolters, F., Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol*, *26*, 183–196.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychol Aging*, *14*, 458–472.

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*, 163–170.

Strauß, A., Kotz, S. A., Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *J Cogn Neurosci*, *25*, 1383–1395.

Strauß, A., Wöstmann, M., Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, *8*, null. doi:10.3389/fnhum.2014.00350.

Sweeney-Reed, C. M., Nasuto, S. J., Vieira, M. F., Andrade, A. O. (2018). Empirical mode decomposition and its extensions applied to EEG analysis: A review. *Adv Data Sci Adapt Anal*, *10*, 1840001.

Taal, C. H., Hendriks, R. C., Heusdens, R., Jensen, J. (2011). An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Trans Audio Speech Lang Process*, *19*, 2125–2136.

Tajima, S., Yamamoto, S., Tanaka, M., Kataoka, Y., Iwase, M., Yoshikawa, E., Okada, H., Onoe, H., Tsukada, H., Kuratsune, H., Ouchi, Y., Watanabe, Y. (2010). Medial orbitofrontal cortex is associated with fatigue sensation. *Neurol Res Int*, *2010*, 671421.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journal Q*, *30*, 415–433.

Tse, C. Y., Lee, C. L., Sullivan, J., Garnsey, S. M., Dell, G. S., Fabiani, M., Gratton, G. (2007). Imaging cortical dynamics of language processing with the event-related optical signal. *Proc Natl Acad Sci U S A*, *104*, 17157–17162.

Uhlmann, R. F., Larson, E. B., Rees, T. S., Koepsell, T. D., Duckert, L. G. (1989). Relationship of hearing impairment to dementia and cognitive dysfunction in older adults. *JAMA*, *261*, 1916–1919.

van den Brink, D., Brown, C. M., Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *J Exp Psychol Learn Mem Cogn*, *32*, 364–372.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang*, *97*, 279–293.

van Wonderen, E., & Nieuwland, M. S. (2023). Lexical prediction does not rationally adapt to prediction error: ERP evidence from pre-nominal articles. *J Mem Lang*, *132*, 104435.

Wambacq, I., Ghanim, I., Greenfield, S., Koehnke, J., Besing, J., Chauvette, C., Yesis, C. (2023). Neurophysiologic patterns of semantic processing of accented speech. *J Neurolinguistics*, *65*, 101117.

Wang, C., Trongnetrpunya, A., Samuel, I. B. H., Ding, M., Kluger, B. M. (2016). Compensatory neural activity in response to cognitive fatigue. *J Neurosci*, *36*, 3919–3924.

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *J Neurosci*, *32*, 14010–14021.

Wilson, R. H., McArdle, R., Watts, K. L., Smith, S. L. (2012a). The Revised Speech Perception in Noise Test (R-SPIN) in a multiple signal-to-noise ratio paradigm. *J Am Acad Audiol*, *23*, 590–605.

Wilson, K. R., O'Rourke, H., Wozniak, L. A., Kostopoulos, E., Marchand, Y., Newman, A. J. (2012b). Changes in N400 topography following intensive speech language therapy for individuals with aphasia. *Brain Lang*, *123*, 94–103.

Winn, M. B., & Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends Hear*, *25*, 23312165211027688.

Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv Adapt Data Anal*, *01*, 1–41.

Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear Hear*, *39*, 293–304.

Yano, M., Suwazono, S., Arao, H., Yasunaga, D., Oishi, H. (2021). Selective adaptation in sentence comprehension: Evidence from event-related brain potentials. *Q J Exp Psychol (Hove)*, *74*, 645–668.

Yu, F., Tan, C. T., Chen, F. (2018). An ERP Study to Evaluate the Quality of Speech Processed by Wiener Filtering. *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*.

Yu, F., Tan, C. T., Chen, F. (2019). Analyzing Auditory Evoked Cortical Response to Noise-suppressed Speech in Cochlear Implant Users Using Mismatch Negativity. *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*.

Zezario, R. E., Fu, S.-W., Chen, F., Fuh, C.-S., Wang, H.-M., Tsao, Y. (2023). Deep Learning-based non-intrusive multi-objective speech assessment model with cross-domain features. *IEEE/ACM Trans Audio Speech Lang Process*, *31*, 54–70.

Zhang, W., Chow, W. Y., Liang, B., Wang, S. (2019). Robust effects of predictability across experimental contexts: Evidence from event-related potentials. *Neuropsychologia*, *134*, 107229.