

Research Article: New Research | Cognition and Behavior

Individual Differences in Cognition and Perception Predict Neural Processing of Speech in Noise for Audiometrically Normal Listeners.

<https://doi.org/10.1523/ENEURO.0381-24.2025>

Received: 29 August 2024

Revised: 27 January 2025

Accepted: 11 February 2025

Copyright © 2025 Shehabi et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

- 1 1. **Manuscript title:** Individual Differences in Cognition and Perception Predict Neural Processing of Speech in
 2 Noise for Audiometrically Normal Listeners.
 3 2. **Abbreviated title:** Cognition Predicts Neural Speech Processing.
 4 3. **Author list:** Sana Shehabi¹, Daniel C. Comstock¹, Kelsey Mankel^{1,2,3}, Brett M. Bormann^{1,4}, Soukhin Das^{1,5},
 5 Hilary Brodie⁶, Doron Sagiv⁶, Lee M. Miller^{1,6,7}.

6 ¹Center for Mind and Brain, University of California, Davis, USA

7 ²Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

8 ³School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

9 ⁴Neuroscience Graduate Group, University of California, Davis, USA

10 ⁵Department of Psychology, University of California, Davis, USA

11 ⁶Department of Otolaryngology | Head and Neck Surgery, University of California, Davis, USA

12 ⁷Department of Neurobiology, Physiology and Behavior, University of California, Davis, USA

- 13 4. **Author contributions:** D.C.C, K.M, H.B, D.S, and L.M.M designed research; S.S, B.B, and S.D performed
 14 research; S.S, D.C.C, K.M, and B.B analyzed data; S.S, D.C.C, K.M, B.B, S.D, D.S, and L.M.M wrote the
 15 paper.
 16 5. Correspondence should be addressed to Dr. Lee M. Miller at leemiller@ucdavis.edu

17

- 18 6. Number of figures: 11 10: Number of words for Significance Statement: 91
 19 7. Number of tables: 3 11: Number of words for Introduction: 750
 20 8. Number of Multimedia: 0 12: Number of words for Discussion: 2,896
 21 9. Number of words for Abstract: 218

22

- 23 10. **Acknowledgements:** This work was supported by the Office of the Assistant Secretary of Defense for
 24 Health Affairs through the Congressionally Directed Medical Research Program (CDMRP) Hearing
 25 Restoration Research Program (HRRP) under Award No. W81XWH-20-1-0485 (to LMM). Opinions,
 26 interpretations, conclusions, and recommendations are those of the author and are not necessarily
 27 endorsed by the Department of Defense. This work was also supported by the Child Family Fund for the
 28 Center for Mind & Brain (to LMM). The authors would like to thank members of the University of California,
 29 Davis Health audiology and clinical research team for performing initial hearing evaluations of our
 30 participants: Dr. Robert Ivory, Au.D., Dr. Mackenzie Quinn, Au.D., Dr. Rachel Krager, Au.D., Dr. Steven
 31 Zurawski, Au.D., Dr. Austin Childers, Au.D., Dr. Kimberly Smith, Au.D., Randev Sandhu, and Angela
 32 Beliveau. We extend our thanks to Cathleen Chan and Jillian McKie for their assistance in data collection.
 33 We are also grateful to Elyse Ehler and Tiana Smith for their efforts in participant recruitment. We would
 34 like to acknowledge Sophie Burstein, Alicia Dye, Reina Itakura, Zachary McNaughton, Ferdous Rahimi,
 35 Tyler Statema, Audrey Vargas, and Nina Wade for their insightful discussions and constructive feedback.
 36 Special thanks to Dr. Chhayakant Patro for providing us with the computerized version of the Trail Making
 37 Test. Finally, we sincerely thank the participants of this research study without whom none of this work
 38 would be possible. This research is published in remembrance of our team member, lab mate, and friend,
 39 Karim Abou Najm.

40

- 41 11. **Conflict of interest:** Lee M. Miller is an inventor on intellectual property related to chirped-speech (Cheech)
42 owned by the Regents of the University of California and not presently licensed.
- 43 12. **Funding sources:** This work was supported by the Office of the Assistant Secretary of Defense for Health
44 Affairs through the Congressionally Directed Medical Research Program (CDMRP) Hearing Restoration
45 Research Program (HRRP) under Award No. W81XWH-20-1-0485 (to LMM). Opinions, interpretations,
46 conclusions, and recommendations are those of the author and are not necessarily endorsed by the
47 Department of Defense. This work was also supported by the Child Family Fund for the Center for Mind &
48 Brain (to LMM).

49 ABSTRACT

50

51 Individuals with normal hearing exhibit considerable variability in their capacity to understand
52 speech in noisy environments. Previous research suggests the cause of this variance may be due to
53 individual differences in cognition and auditory perception. To investigate the impact of cognitive and
54 perceptual differences on speech comprehension, 25 adult human participants with normal hearing
55 completed numerous cognitive and psychoacoustic tasks including the Flanker, Stroop, Trail Making,
56 Reading Span, and temporal fine structure (TFS) tests. They also completed a continuous multi-talker
57 spatial attention task while neural activity was recorded using electroencephalography (EEG). The
58 auditory cortical N1 response was extracted as a measure of neural speech encoding during continuous
59 speech listening using an engineered “chirped-speech” (Cheech) stimulus. We compared N1
60 component morphologies of target and masker speech stimuli to assess neural correlates of attentional
61 gains while listening to concurrently played short story narratives. Performance on cognitive and
62 psychoacoustic tasks were used to predict N1 component amplitude differences between attended and
63 unattended speech using multiple regression. Results show inhibitory control and working memory
64 abilities can predict N1 amplitude differences between the target and masker stories. Interestingly, none
65 of the cognitive and psychoacoustic predictors correlated with behavioral speech-in-noise listening
66 performance in the attention task, suggesting that neural measures may capture different aspects of
67 cognitive and auditory processing compared to behavioral measures alone.

68

69 SIGNIFICANCE STATEMENT

70

71 These findings contribute to our understanding of how cognition affects the neural encoding of
72 auditory selective attention during speech perception. Specifically, our results highlight the complex
73 interplay between cognitive abilities and neural encoding of speech in challenging listening
74 environments with multiple speakers. By incorporating these additional measures of cognition, we can
75 achieve a more comprehensive understanding of an individual’s speech perception abilities, even in
76 individuals with normal hearing. This approach could lead to earlier detection of hearing issues and
77 more personalized interventions, ultimately enhancing communication outcomes for those with hearing
78 difficulty.

79

INTRODUCTION

Speech perception is a complex and multileveled process that requires the coordination of numerous neural systems, from decoding basic acoustic features of speech to transforming them into meaningful linguistic representations. Comprehension in noisy environments presents an additional challenge, placing substantial demands on every level of processing. This includes not only the encoding of stimuli and linguistic representations but also the segregation of auditory information into distinct streams (Griffiths & Warren, 2004), the ability to focus on a specific talker while filtering out others (Alain & Arnott, 2000; Shinn-Cunningham, 2008), and maintenance information over time (Martin, 2021). Consequently, speech-in-noise (SiN) performance is linked to both auditory and domain-general cognitive mechanisms (Akeroyd, 2008; Holder et al., 2018; Zekveld et al., 2013).

SiN perception disproportionately challenges listeners with hearing loss. However, many individuals with normal hearing thresholds also struggle with speech comprehension in such conditions. In fact, as many as 15% of individuals seeking hearing assistance at audiology clinics have normal hearing thresholds but identify difficulty understanding speech in noisy settings as their primary concern (Cooper & Gates, 1991; Hind et al., 2011; Tremblay et al., 2015). Even among listeners with normal thresholds, significant variability in speech-in-noise recognition can still be observed (e.g., Ruggles & Shinn-Cunningham, 2011).

This variability in SiN understanding could be influenced by individual differences in cognition and general perceptual abilities. Previous work suggests the most critical cognitive mechanisms for SiN performance include selective attention (Shinn-Cunningham, 2008), working memory (Akeroyd, 2008; Gordon-Salant & Fitzgibbons, 1997; Ronnberg et al., 2008), inhibitory control (Janse, 2012, Sommers & Danielson, 1999), and executive function (Perrone-Bertolotti et al., 2017). Deficiencies in these domains can significantly impair speech perception (Pichora-Fuller et al., 1995). In addition to cognition, non-audiometric psychoacoustic abilities are also important, particularly sensitivity to temporal fine structure (TFS). TFS sensitivity is essential for spatial hearing, auditory object segregation and streaming, listening in the gaps, and pitch perception (Moore, 2008). As a result, TFS sensitivity is crucial for analyzing complex, ecologically valid auditory scenes, especially when speech is presented with background noise (Füllgrabe et al., 2015; Oberfeld & Klöckner-Nowotny, 2016). Notably, Oberfeld & Klöckner-Nowotny (2016) found that both cognitive measures, including selective attention, and psychoacoustic measures, such as TFS sensitivity, can predict SiN performance in individuals with normal hearing. Thus, explorations of speech-in-noise recognition abilities must also account for individual differences in both cognitive and basic auditory skills.

Additionally, successful SiN recognition relies on robust neural encoding mechanisms within the auditory system. The N1 component, part of the P1-N1-P2 complex, reflects cortical processing of sounds (Winkler et al., 2013). The N1 is considered an index of early perceptual encoding of speech sounds, and it is sensitive to acoustic cues within a stimulus, speech or otherwise (Getz & Toscano, 2021). N1 amplitude has been shown to reflect allocation of attentional resources to specific sounds, and it is sensitive to attentional modulation in multi-talker listening scenarios (Kerlin et al., 2010; Picton & Hillyard, 1974; Stapells, 2009). Previous research suggests that N1 amplitude is larger for attended sounds compared to ignored ones, highlighting its role in auditory selective attention (Hillyard et al.,

121 1973) and segregation of competing sound sources (Gutschalk et al., 2007; Snyder et al., 2006).
122 Presumably, individual differences in speech perception should also be linked with variability of neural
123 processes within the auditory cortex. Yet, despite the established influence of cognitive and non-
124 audiometric psychoacoustic abilities on speech perception, how these individual differences impact the
125 neural encoding of speech (and SiN; Bednar & Lalor, 2020) remains unclear (for review, see Davidson
126 & Souza, 2024). By investigating N1 amplitude differences for target and masker speech, our study
127 aims to better characterize the variability in speech perception abilities among normal-hearing
128 individuals in noisy environments.

129 The goal of this study is to understand how cognitive factors influence speech processing in the
130 auditory cortex. We focus on N1 amplitude differences between target and masker speech to
131 understand the brain's mechanisms for prioritizing relevant auditory information over irrelevant stimuli.
132 We hypothesize that neural mechanisms involved in cocktail party listening are shaped not only by
133 basic auditory factors, such as hearing sensitivity, but also by an individual's cognitive abilities.
134 Specifically, we expect individuals who perform better on cognitive and psychoacoustic tasks to show
135 larger N1 amplitude differences between target and masker speech. Enhanced cognitive and perceptual
136 abilities are expected to improve auditory stream segregation, resulting in greater differentiation
137 between the two speech streams. Collectively, these findings shed light on the cognitive contributions
138 to neural mechanisms relevant for auditory selective attention during SiN perception.

139

140

141

METHODS

Participant Information

142
143 Twenty-nine participants were recruited for this study. Eligible participants were required to be
144 between 18 and 40 years old, speak English as their first language, and report no neurological,
145 psychiatric, or scalp conditions that could directly impair their ability to understand or attend to speech
146 or hinder electrophysiological measurements (e.g., epilepsy, certain strokes, ADHD, scalp issues such
147 as wounds). Participants were screened for visual acuity (< 20/40 per Snellen chart), normal hearing
148 (<25 dB HL air conduction thresholds between 250-8000 Hz in both ears, no air-bone gaps >10 dB, no
149 interaural asymmetries >20 dB at 500, 1000, or 2000 Hz), and cognitive abilities (>24 MoCA score)
150 (Nasreddine et al., 2005).

151 One participant withdrew from the study due to personal reasons, two were excluded for failing
152 the hearing screening at the time of testing, and one was excluded due to technical issues with data
153 acquisition. Additionally, after fitting the model, one participant was identified as an influential outlier in
154 the Trail Making Test using a Cook's Distance test threshold of 0.5 based on the multiple regression
155 analysis (see Statistical Analysis section below). This resulted in a final cohort of 24 participants. The
156 average age of participants was 23.4 ± 5.02 SD years (range: 18-38 years). Fourteen participants were
157 female and ten were male. Twenty-one participants self-identified as right-handed, one as left-handed,
158 and two as ambidextrous. On average, participants had 17.00 ± 3.18 SD years of education and 4.21
159 ± 3.82 SD years of formal music training. Pure tone averages (PTA) were obtained for each ear as the
160 average threshold at 0.5, 1, and 2 kHz. Average air conduction PTA thresholds were 6.88 ± 3.78 SD
161 and 7.04 ± 4.56 SD for the right and left ears, respectively. All participants provided informed consent

162 as approved by the university's Institutional Review Board and were financially compensated for their
163 participation and travel.

164

165 **Project Overview**

166 Each participant completed three sessions in total: a 1-hour audiological exam, a 1.5-hour
167 behavioral session, and a 2.5-hour laboratory EEG session. Most participants completed more than
168 one session on the same day (e.g., audiological + behavioral or behavioral + EEG), with a long break
169 in between to minimize fatigue. The analysis presented in this paper is based on a subset of data from
170 a larger study investigating hearing across the lifespan, which will be detailed in future reports.
171 Consequently, only a portion of the measurements and tests conducted during the study are included
172 in this analysis.

173

174 **Audiology Session**

175 Participants completed a standard audiologic assessment either conducted by a professional
176 audiologist or by highly trained personnel in the lab. Before inserting earphones, an otoscopic
177 examination was conducted to check for tympanic membrane pathologies or any obstructions in the ear
178 canal (i.e., cerumen) that could prohibit the use of insert earphones. Participants found to have ear
179 canal occlusions or other visible pathologies were advised to consult their primary care provider and
180 return to the study once the issue was resolved.

181 Pure-tone audiometry was performed for both ears either at a local affiliated audiology clinic or
182 in a quiet, sound-dampened room within the lab. Air-conduction thresholds were measured using pulsed
183 pure tones at frequencies of 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz with insert earphones. Bone-conduction
184 thresholds were measured using a RadioEar bone conductor placed on the mastoid at 0.25, 0.5, 1, 2,
185 3, and 4 kHz. The thresholds were determined using a 10 dB down/5 dB up staircase procedure. Pure
186 tone averages (PTA) were obtained for each ear as the average air-conduction threshold at 0.5, 1, and
187 2 kHz. While several other tasks typical of a standard audiology exam were conducted, only air and
188 bone conduction measures are described here as they were the indicators used to determine normal
189 hearing in participants.

190

191 **Cognitive Tasks**

192 Cognitive tasks were administered in a randomized order for each participant.

193

194 *1. Flanker Test*

195 The Flanker test measures visuospatial interference, selective attention, and inhibitory control (Zelazo
196 et al., 2013). We used a computerized version of the task using the PEBL software (Psychology

197 Experiment Building Language; Mueller & Piper, 2014). In the task, five light gray arrows were displayed
198 in the center of the screen against a black background. The arrows were either all pointing in the same
199 direction (congruent condition) or with the center arrow pointing in the opposite direction to the flanking
200 arrows (incongruent condition). Participants were asked to press the left or right shift keys on a keyboard
201 corresponding to the direction of the center arrow. A fixation cross appeared for 500 ms before the
202 arrows were presented, and participants had 2 seconds to respond before the trial timed out. A blank
203 screen was shown for 2 seconds before the next trial began. Participants completed 8 practice trials,
204 with two trials for each combination of conditions (left/right, incongruent/congruent), followed by 80 test
205 trials (20 per condition). The interference effect was calculated as the median reaction time difference
206 between incongruent and congruent trials.

207

208 *2. Stroop*

209 The Stroop test is a long-standing, well-established assessment of linguistic interference and
210 response inhibition (Strauss et al., 2006; Stroop, 1935). In our study, we utilized a computerized version
211 of the Victoria Stroop Test (VST; Strauss et al., 2006), implemented using PEBL. The VST was selected
212 for our study due to its shorter test administration time compared to other standardized Stroop tests.
213 Evidence suggests that shorter test durations may be more sensitive to individual differences on this
214 task (Klein et al., 1997). The VST has been proven effective with a wide range of adults, from 18 to 94
215 years old (Troyer et al., 2006).

216 As with other Stroop versions, the VST involves color identification of items presented to the
217 participant, including a color word interference task. The VST is divided into three blocks always
218 presented in the following order: colored dots (Part D), neutral words (Part W), and color words
219 presented in contrasting text colors (Part C). Each block displays 24 items in a 6 x 4 grid on a gray
220 background. Items are colored red, blue, green, or yellow, and participants press a corresponding
221 number key to identify the color of each item. A square is shown around the current item trial, and it
222 only moves to the next item with a correct response. The square flashes white when an incorrect
223 response is made. Each color is used six times, and the item colors and the color-number keypad pairs
224 are pseudorandomized for every participant (one of each color in each row). Part D serves as a practice
225 block for color-number mapping and is not scored here. For Part W (neutral words) and Part C (color
226 words), participants must name the colors in which the words are printed, disregarding the verbal
227 content printed in lowercase letters (neutral words: when, hard, and, over; color words: red, green, blue,
228 yellow). In Part C, the color name never corresponds to the text color, thus requiring individuals to inhibit
229 an automatic reading response and only identify the text color itself. The interference effect is measured
230 as the difference in median reaction times between Part C (i.e., the linguistic interference task) and Part
231 W (i.e., the control task).

232

233 *3. Reading span*

234 Working memory ability was assessed using the reading span test (Daneman & Carpenter,
235 1980), as employed in previous research (Dryden et al., 2017; Lunner, 2003; Rudner et al., 2012). We

236 administered a computerized version of the reading span via the PEBL program. The PEBL reading
237 span test was developed for automated scoring as described by Unsworth et al. (2009). In this task,
238 participants read sentences displayed on the computer screen and indicated whether each sentence
239 made sense by pressing a key (i.e., in grammatical, coherent, or plausibility sense). Sentence lengths
240 ranged between 9 and 14 words. After the sentence, a letter was presented on the screen for 1 second,
241 and participants were required to memorize the sequence of letters presented across sentence-letter
242 pairs. The task consisted of blocks of 3 to 7 sentence-letter pairs, with each block length presented
243 three times in random order throughout the test. At the end of each block, participants selected the
244 correct sequence of letters from a 4 x 3 grid displayed on the screen and were instructed they could
245 guess or leave a blank in the letter sequence if they were uncertain. They were then provided feedback
246 regarding both sentence judgment and letter sequence accuracy.

247 Prior to the test, participants completed two practice examples each for sentence reading, letter
248 memorization, and sentence-letter memorization pairs. The average reading time during these practice
249 trials determined the timeout duration for the test blocks (i.e., $\text{timeout} = 2.5 \times \text{average practice reading}$
250 time) (Unsworth et al., 2009). If participants did not respond to a sentence before the timeout period,
251 the sentence trial was marked incorrect, and the subsequent letter was shown on the screen.
252 Throughout the test, sentences, letters, and feedback were shown in white text on a black screen
253 background.

254 Working memory span was determined by calculating the average accuracy of letter recall for 3
255 and 4 letter trials, when any letters were correctly recalled in the correct order (i.e., each correct letter
256 contributing to the score). Though earlier research has suggested a working memory capacity of up to
257 ~7 items (Miller, 1956), recent studies indicate that this capacity is lower and can vary based on task
258 demands and circumstances (Cowan, 2015). Consistent with findings indicating working memory
259 capacity is approximately 4 items (for reviews, see Cowan, 2001, 2015), our findings revealed that the
260 accuracy for the 3 and 4 letter trials provided more stable and informative data for our model compared
261 to the average accuracy across 3 to 7 letter trials (pilot data not shown). Therefore, we used the average
262 accuracy from the 3 and 4 letter trials as our primary measure of working memory span.

263

264 4. Trail making test

265 The Trail Making Test (TMT) (Reitan, 1958) was used to assess cognitive flexibility and
266 executive function. In TMT-A, participants connect lines sequentially between numbers, while TMT-B
267 requires them to alternate between numbers and letters (e.g., 1-A-2-B). Participants were given TMT-
268 A followed by TMT-B, with a short practice at the beginning of each. The test was administered via
269 PEBL, with white nodes displayed on a gray background and black text for letters and numbers.
270 Participants clicked on the nodes in the correct sequence as quickly and accurately as possible, and
271 lines were drawn on the screen to connect consecutive nodes upon correct selection. Given known
272 limitations of the original Reitan-developed TMT A/B (i.e., length differences between the nodes of the
273 A and B versions) (Gaudino et al., 1995), our version includes fixed node positions as mirrored images
274 across TMT-A and B to ensure similar distance lengths for the two tests (Strelcyk et al., 2019). Any
275 differences in reaction time are therefore presumed to reflect processing differences across the A and

276 B versions. Executive function and cognitive flexibility were measured as the difference in median
277 reaction times of TMT-B and TMT-A conditions.

278

279 *5. Temporal fine structure*

280 The ability to localize low-frequency sounds along the horizontal plane relies on interaural time
281 differences (ITD), which are important cues for distinguishing multiple sounds such as recognizing
282 speech from a target talker amidst spatially separated background noise or other interfering sounds
283 (Moore, 2021). For sinusoidal signals, ITDs are equivalent to interaural phase differences (IPDs), and
284 the ability to discriminate changes in IPDs depends on binaural sensitivity to temporal fine structure
285 (TFS). To assess IPD discrimination abilities, we used the TFS-LF (low-frequency) test (Hopkins &
286 Moore, 2010), which is implemented in a computerized software program developed by Søk and Moore
287 (Søk & Moore, 2012, 2021). In the TFS-LF test, participants are tasked with discriminating sinusoidal
288 tone bursts with adaptively varying IPDs, where the tone envelopes are synchronized across both ears,
289 ensuring that task performance relies on sensitivity to TFS (IPD). TFS-LF administration was consistent
290 with previous work and is briefly described below (Füllgrabe & Moore, 2018; Hopkins & Moore, 2010;
291 Søk & Moore, 2012, 2021).

292 IPD discrimination was evaluated at test frequencies of 250 and 500 Hz, consistent with previous
293 studies (for review, see Füllgrabe & Moore, 2018). Intensity level was set individually for each ear 30
294 dB above the participant's pure tone thresholds at the tested frequencies (i.e., 30 dB sensation level
295 [SL] at 250 and 500 Hz). The TFS-LF task consisted of a two-interval, two-alternative-forced-choice
296 design, where each interval included four successive tones. In one of the intervals, all four tones had a
297 consistent 0° IPD, while in the other interval, the IPD alternated between 0° and ϕ across the four
298 successive tones. A 0° IPD is perceived as a tone localized to the center of the head, whereas a large
299 IPD is perceived as a tone lateralized (or as more diffuse) toward one ear. Participants were asked to
300 identify the interval in which the tones sounded different or appeared to change in some way. Each tone
301 had a duration of 400 ms, with 20 ms rise and decay times and a 300 ms inter-interval gap between
302 tone blocks. Feedback was provided after each trial. The IPD ϕ was adaptively varied, starting from
303 180° (maximally lateralized), using a 2-down, 1-up procedure to estimate a threshold corresponding to
304 71% correct responses. That is, the value of ϕ is divided by a factor k after two successive correct
305 responses or multiplied by factor k following one incorrect response. Before the first turn point, $k =$
306 1.253; between the first and second turn points, $k = 1.252$; and $k = 1.25$ after the second turn point. The
307 threshold was computed as the geometric mean value of ϕ at the last six turn points, and the final TFS-
308 LF measure was then calculated as the average across both 250 and 500 Hz frequencies.

309

310 *Additional tests*

311 During the cognitive session, participants also completed two additional tests: 1) a pitch
312 discrimination task and 2) a Speech Recognition in Noise Test (SPRINT), which assesses SiN
313 perception. These measures were not included as predictors in the current analysis, as our primary

314 focus was on evaluating variables that most closely align with existing research on predictors of SiN
315 perception (Akeroyd, 2008; Janse, 2012; Perrone-Bertolotti et al., 2017; Shinn-Cunningham, 2008).

316

317 **Continuous Multi-Talker Spatial Attention Task**

318 To evaluate SiN perception abilities, participants completed a continuous multi-talker spatial
319 attention task.

320

321 *Stimuli*

322 The stimuli consisted of short fairytale stories available in the public domain (available for
323 listening on www.librivox.com). Stories were selected to have content that was unlikely to be widely
324 recognized, reducing the chances of participants relying on modern renditions for comprehension recall
325 (cf. stories popularized by children’s books or movies). Additionally, each story was chosen to have a
326 similar narrative arc—i.e., up through rising action, climax, or early falling action— within a 7.5-minute
327 block to ensure sufficient and consistent content for probing comprehension. Thirteen stories were
328 piloted with four raters who evaluated them based on factors such as interest and engagement, ease
329 of understanding, complexity of grammar or writing style, familiarity (i.e., low scores preferred), and
330 suitability for an adult audience (not only for children). The seven highest-scoring stories across these
331 criteria were selected for use in the experiment.

332 The short stories were revised to incorporate monosyllabic color words as targets for the
333 experiment (e.g., red, green, black, white, etc.). The color words were integrated into the story itself
334 rather than being inserted as unrelated, random target words (e.g., “...with a narrow *blue* ribbon over
335 her shoulders”). A male native English speaker with an American Midwestern accent recorded the
336 modified stories using a Shure KSM244 vocal microphone (cardioid polar pattern, high-pass filter at 18
337 dB per octave with an 80 Hz cutoff) and Adobe Audition (sample rate = 48,000 Hz, 32-bit depth [float])
338 in a soundproof, quiet room. Silent gaps in the recordings longer than 500 ms were shortened to 500
339 ms, following procedures similar to those in previous studies (Broderick et al., 2018; Teoh et al., 2022;
340 Teoh & Lalor, 2019). Finally, the edited recordings were cropped to 7.5 minutes per story.

341

342 *Cheech*

343 The short story auditory stimuli were subsequently modified using a patented technique known
344 as “Cheech” (“chirped-speech”), which is intended to elicit robust auditory evoked potentials from
345 brainstem to cortex (Backer et al., 2019; Miller et al., 2020). In addition to providing insight into multiple
346 levels of neural processing simultaneously (e.g., ABR, MLR, LLR), Cheech provides greater acoustic
347 control over the specific acoustic parameters in naturalistic speech used to evoke the brain responses.
348 We describe the details in this paper for completeness, but some of the details may not be relevant for
349 LLRs. In Cheech, some of the glottal pulse energy within specified frequency bands is replaced with
350 narrowband synthetic chirps (for more details, see Backer et al., 2019). These chirps are aligned with

351 the natural glottal pulse energy in voiced speech, creating an acoustically fused auditory perceptual
352 object. This modification maintains the naturalistic qualities and linguistic content of speech while
353 introducing sufficient transient chirp activation to measure auditory brainstem, thalamic, and cortical
354 responses.

355 The process begins by identifying voiced epochs where chirps will be inserted. Audio was filtered
356 from 20 to 1000 Hz, and voiced periods of at least 50 ms were defined when the speech envelope
357 between 20-40 Hz exceeded a threshold of approximately 28% of the overall speech root-mean-square
358 (RMS) amplitude. The timing of the glottal pulses within these voiced periods was determined through
359 a speech resynthesis process using custom MATLAB code (The MathWorks Inc., 2021) and the
360 TANDEM-STRAIGHT toolbox, which retains the original speech periodicity characteristics (Kawahara
361 et al., 1999, 2008; Kawahara & Morise, 2011). The continuous speech was then re-filtered into
362 alternating, octave-wide frequency bands: 0-250, 500-1000, 2000-4000, and 11,000-∞ Hz, with chirp
363 energy constrained to frequency bands of 250-500, 1000-2000, and 4000-11,000 Hz. The chirp and
364 speech bands in the alternating, interleaved octave-wide bands were then added together. Each mono
365 track was then duplicated to form stereo audio. Before the Cheech synthesis, audio files were
366 normalized to ensure consistent chirp amplitudes relative to the speech levels across each story, and
367 the RMS amplitudes were re-checked across all Cheech-modified stories to ensure equivalence after
368 the Cheech process was completed.

369 The timing of the chirps varied with the natural fluctuations in the running speech, with special
370 modifications to optimize measurement of auditory event-related potentials (AERPs). Specifically, when
371 voicing energy was detected, chirps were inserted at minimum 18.2 ms apart (55 Hz), skipping glottal
372 pulses that occur within the 18.2 ms window. The first chirp in any sequence was always followed by a
373 minimum interval of 48 ms before the next chirp was presented, skipping glottal pulses in between. This
374 longer ISI was designed to enhance the measurement of middle- latency responses (MLRs); we use
375 this “first chirp” to elicit LLR N1s as well.

376 In the dual-talker conditions, one story from each pair was pitch and vocal-tract modulated using
377 MATLAB STRAIGHT prior to the Cheech process. Specifically, stories intended to represent a female
378 voice were pitch modulated to an average fundamental frequency (F0) of 180 Hz, while the original
379 male voice audio had an average F0 of approximately 128 Hz. By modulating the F0, key speaker
380 characteristics such as phrasing, intonation, and pacing were maintained, as all stories were initially
381 recorded by a single talker. Additionally, this pitch modification created a perceptual distinction between
382 the two voices, introducing a combined voice-gender and spatial release from masking effect. This
383 effect, which is consistent with prior findings (Oh et al., 2021, 2022), facilitated better differentiation of
384 the spatially separated voices and improved task performance during pilot testing.

385

386 *Continuous multi-talker spatial attention task*

387 The Cheech stimuli were spatially filtered using head-related transfer functions (HRTFs) from
388 the SADIE II database (Armstrong et al., 2018) to simulate audio at $\pm 15^\circ$. This spatialized audio was
389 mirrored by two symbols on two computer screens centered approximately 15° to the left and right of

390 the participant's midline. The symbols "<" and ">" indicated the target story's location on the left or right,
391 respectively, while a "+" symbol denoted the location of the target story, as shown in Figure 1. The
392 visual symbols were identical in size, color, and line lengths to ensure consistent luminance for eye-
393 tracking measurements (not analyzed in this report). Custom MATLAB Psychtoolbox code (Brainard,
394 1997; Pelli, 1997) was used to program the spatial and visual stimuli to switch between left and right
395 locations 75 times over a 7.5-minute period, with each switch occurring approximately every six seconds
396 (random distribution of 6 ± 1 seconds). The visual icons switched instantaneously at each switch cue,
397 while the audio crossover fade time was set to 35 ms, starting at the onset of each switch cue. During
398 piloting, this fade duration was found to effectively minimize audio artifacts, such as "pops," while also
399 reducing the perception of a gradual audio glide between spatial locations (i.e., as sudden of a jump as
400 possible without introducing noticeable distortions).

401

402

[FIGURE 1]

403

404 75 switches were placed in the story, occurring every 6 seconds on average. 25 of those
405 switches were specifically timed to precede the color words. The time intervals between these switches
406 and the color words were 0.125, 0.25, 0.5, 1, and 2 seconds, with the intervals randomly assigned to
407 the color words so that each interval duration occurred exactly 5 times. These intervals were determined
408 based on preliminary piloting that aimed to evaluate whether brief lapses in attention following a shift in
409 spatial focus—analogue to "attentional blinks"—affected participants' ability to detect color words;
410 these results will be presented in a subsequent paper. Additionally, the 25 color words were evenly
411 distributed between left-to-right and right-to-left switches within each story, minus the one remainder.
412 Onset timings for the color words were calculated using Gentle, an audio forced aligner
413 (<https://github.com/lowerquality/gentle>). The Gentle forced aligner software finds the start and end
414 points of spoken words, allowing us to pinpoint exactly when the color words start, which was needed
415 for determining switch times.

416

417 *Experimental conditions*

418 After cropping the audio, the seven stories were assigned to specific experimental conditions:
419 two stories presented alone ("mono-talker"), two sets of paired stories ("dual-talker"), and one for initial
420 practice. Half of the stories (one mono-talker and one dual-talker pair) had modified Cheech
421 characteristics to study their effects on brainstem encoding (results forthcoming in future reports);
422 however, this modification did not impact cortical responses, so we used all six non-practice stories for
423 our N1 analyses. Since our focus is on predicting SiN performance, this paper primarily examines the
424 dual-talker conditions (excluding Figure 4 and associated analyses). In the dual-talker paradigm, each
425 story was presented twice—once as the "target story" and once as the "masker story" within its pair.
426 This experimental manipulation allowed us to assess the effects of directed spatial attention, as
427 participants were instructed to either "attend" to or "ignore" the same audio depending on its role in the
428 trial. The order of stories and conditions was counterbalanced across participants using a Latin square

429 design to minimize potential sequence-dependent learning effects. Additionally, no pair of stories was
430 presented consecutively.

431 Story pairs in the dual-talker condition required additional considerations. As described in the
432 Cheech section above, one voice in each story pair was pitch and vocal tract modulated to simulate a
433 female speaker. Consistent with findings on the combined effects of voice-gender and spatial release
434 from masking (Oh et al., 2021, 2022), two distinct talker F0s facilitated better listener differentiation
435 between spatially separated voices and improved task performance during piloting. The mono-talker
436 stories were presented in their original, unmodified male voice. To minimize the influence of stimulus
437 variability on the N1, the analysis in this paper focused solely on the male target and male masker
438 conditions. Due to the naturalistic embedding of color words within the stories, some overlap occurred
439 in the dual-talker pairings, i.e., color words in the target and masker speech occurred in close temporal
440 proximity. Story pairs were selected to minimize color word overlaps, resulting in six overlapping words
441 that were excluded from the behavioral analysis (four from one story pair and two from the other pair).
442 A few target story switch times were adjusted to ensure that no color words in the masker story occurred
443 during the switch (i.e., masker color word onset was not between -0.5 to 0.1 ms from a switch time).

444

445 *Stimulus presentation equipment*

446 Participants were seated in a soundproof, electromagnetically shielded booth approximately 176
447 cm from two computer monitors. The subjects and cables were positioned at a sufficient distance from
448 the monitors. Testing confirmed that this arrangement effectively prevented any electromagnetic
449 interference from the monitors. The Dell UltraSharp monitors had a 60 Hz refresh rate with screen
450 settings maintained at standard factory defaults. Stimulus presentation, including both audio and visual
451 components, was controlled using custom MATLAB and Psychtoolbox code. Auditory stimuli were
452 routed from a Hammerfall audio card to an RME Fireface UFX II audio interface, then through an RME
453 ADI-2DAC FS headphone amplifier, and finally delivered via ER-2 insert earphones (Etymotic
454 Research/Interacoustics). The ER-2 insert wires were electromagnetically shielded up to the transducer
455 box to minimize EEG artifacts (Campbell et al., 2012). The Cheech stimuli were presented at 75 dB
456 SPL.

457 In addition to the 2-channel audio tracks, a pair of channel outputs carrying trigger events was
458 routed from the Fireface audio interface to two Brain Products StimTraks. These StimTraks monitored
459 triggers for the target and masker stories, respectively, converting the audio trigger signals into pulses
460 sent to a TriggerBox (Brain Products GmbH). This configuration ensured that audio playback was
461 synchronized with the chirp triggers embedded during the Cheech process, minimizing latency jitter,
462 and maintaining sample-level synchronicity. Triggers managed by Psychtoolbox—such as participant
463 responses, block sequence triggers, and spatial location switches—were also routed to the TriggerBox
464 through the computer parallel port. The TriggerBox then transmitted both parallel port and StimTrak
465 triggers to the Biosemi signal receiver box (refer to EEG recordings). To calibrate our stimuli, we used
466 a Larson-Davis PRM902 preamplifier connected to a Larson-Davis 2221 preamplifier power supply,
467 paired with a B&K Ear Simulator Type 4157 (IEC 60318-4 Coupler). Audio stimuli were calibrated using
468 either C-weighting (for Cheech stimuli) or Z-weighting (for cognitive and speech-in-noise tests), with a

469 default preamplifier gain of 20 dB applied. Signal intensity measurements were then analyzed using a
470 custom MATLAB script to ensure accuracy and consistency across all stimuli.

471

472 *Attention task procedure*

473 The order of events within each block is as follows: a 1-minute rest period (for resting state EEG
474 analysis), 7.5 minutes of audio presentation, comprehension questions, subjective workload questions,
475 a flash reaction time task, and a break period. During the rest period, participants were instructed to
476 remain still with their eyes open. They were instructed to attend to the target story as the auditory stream
477 alternates between left and right spatial locations. They must immediately shift both their selective
478 attention and eye gaze to the fixation point (“+”) on the computer screen as it switches the cue.
479 Participants were also instructed to press the spacebar upon hearing the embedded color words in the
480 story as quickly and accurately as possible. In dual-talker condition blocks, they were instructed to
481 ignore color words from the masker story and only respond to those from the target story. Following the
482 audio presentation, participants completed various behavioral tasks, including comprehension
483 questions, subjective task workload assessments, and reaction time tests which are not discussed in
484 this paper. Finally, participants were given a break before starting the next block. The entire experiment
485 presentation was controlled by custom MATLAB and Psychtoolbox code.

486

487 *Measure of SiN Perception*

488 SiN perception during listening was measured by the correct identification of color words
489 embedded in the target story, referred to as “hits”. A hit was defined as pressing a designated keyboard
490 button within 2 seconds after the onset of a target word. Any color words in the dual-talker condition
491 that occurred within ± 2 seconds of each other in the target and masker were excluded from analyses.
492 Specifically, the relevant measure was the proportion of correct hits out of the total possible, non-
493 overlapping target words (n=25 in the mono-talker condition, n=21 in the dual-talker condition).

494

495 **EEG**

496 *EEG recordings*

497 EEG recordings were obtained in an electrically shielded booth using a BioSemi ActiveTwo
498 system (BioSemi B.V., Netherlands, www.biosemi.com) with a sample rate of 8192 Hz from 64
499 electrodes following the International 10-20 standard, plus an additional 4 electrode pairs around ears
500 (earlobe, mastoid, posterior to the earlobe and inferior to the mastoids, and superior-anterior to the
501 tragus near the zygomatic arch) for 72 total electrodes. Prior to recording, offsets from all electrodes
502 were measured to be below 20 μ V relative to the Common Mode Sense electrode using ActiveView2
503 software. Data were recorded using Lab Streaming Layer (LSL) software
504 (<https://github.com/sccn/labstreaminglayer>) in XDF format on a desktop computer running Windows
505 10. Events were recorded alongside EEG data as a separate trigger channel in the XDF file.

506

507 *EEG Preprocessing*

508

509 The EEG data were preprocessed in MATLAB 2021b using EEGLAB (Delorme & Makeig, 2004)
510 and custom MATLAB code. Data were first imported from XDF files into EEGLAB using the XDF
511 importer plugin (version 1.19 <https://github.com/xdm-modules/xdm-EEGLAB>) after which events were
512 extracted from the trigger channel using custom MATLAB code. Given the high number of densely
513 packed event codes (>130,000/subject at an average rate of 1 event every 20ms), events from the 3
514 different event streams (parallel port and 2 StimTraks) often overlapped in the trigger channel, causing
515 occasional event code errors. We therefore checked all recorded events against expected events and
516 corrected for any missing, extra, or jittered events using custom MATLAB code. We then applied a
517 second set of corrections to adjust for differences in the acoustic signal and the event times. These
518 corrections accounted for the 1.973 ms audio delay due to the HRTF process, a 1 ms audio delay due
519 to the audio travel time from the transducer via the insert earphone tubes, a 1 ms trigger delay in one
520 of the StimTrak channels caused by the TriggerBox hardware, and a 0.541 ms trigger delay to the chirp
521 triggers accounting for a difference between the expected chirp timing from the Cheech process and
522 the chirp onsets.

522

523 Data were then pruned to retain only sections of the data from the rest periods and the attention
524 task periods with 10 second buffers on both sides of the selected periods to prevent filtering artifacts.
525 Each remaining section was then DC corrected and filtered using a non-causal 2nd order 0.1 Hz high-
526 pass Butterworth filter using ERPLAB's filter function to remove slow drifts and then referenced to
527 electrode Cz. Data were then checked and cleaned for 60 Hz line noise, which was occasionally found
528 in some channels despite the electrically shielded booth. Only channels with detected line noise were
529 cleaned (mean = 4.5 channels \pm 7.7 SD, range for a given subject = 0-39 channels) to minimize the
530 effects of the cleaning. This was achieved by first checking for line noise using a frequency tagging
531 approach. Individual channel spectra were calculated across all frequencies. For each frequency, x , the
532 average spectral power of 4 neighboring frequencies except those immediately adjacent ($x-2$, $x-3$, $x+2$,
533 $x+3$) were subtracted. If the resulting neighbor corrected power at 60 Hz or any of its first 4 harmonics
534 (120, 180, 240, 300 Hz) was greater than 8 standard deviations calculated from the corrected power
535 from all other frequencies, that channel was cleaned of line noise using the Cleanline plugin
536 (<https://www.nitrc.org/projects/cleanline>). This repeated up to a maximum of 5 times, or until no further
537 line noise was detected.

537

538 We used Independent Component Analysis (ICA) to further clean the data of eye blink, eye
539 movement, and heartbeat artifacts following the methodology outlined by Luck (2022). This involved
540 making a copy of the data processed up to this point (ICA copy), down sampling and cleaning the ICA
541 copy, followed by ICA and marking and rejecting artifact components, then transferring the cleaned ICA
542 weight matrix from the ICA copy to the original data. This approach allows for a massive reduction in
543 ICA computation time, while still retaining all the information in the data relevant for decomposing eye
544 and heart artifacts. Specifically, after making a copy for ICA, we down-sampled to 256 Hz to speed ICA
545 processing, and then manually inspected and rejected bad channels (mean=3.67 \pm 2.16 SD channels,
546 range=0-9 channels). Following channel rejection, a non-causal 1 to 50 Hz 8th order Butterworth
547 bandpass filter using ERPLAB's filter function was applied. We then removed sections of data with
548 significant movement and muscle artifacts that can hinder ICA's ability to isolate eye and heart

548 components, using ERPLAB's continuous artifact rejection function (mean=126.54 \pm 86.85s SD or
549 3.52% \pm 2.36%, range=4.98-298.79s or 0.14-7.96%). The rejection process operated with a 1 s sliding
550 window with a 0.25 s step size and an initial rejection threshold of \pm 200 μ V. In order to retain eye activity
551 for ICA, the rejection threshold was applied against all channels except those mostly likely to contain
552 eye related activations (FP1, FPz, FP2, AF7, AF3, AFz, AF4, AF8). The rejection threshold was
553 inspected and adjusted for each subject to maximize rejecting movement and muscle artifacts while
554 retaining eye blinks (mean=220.83 \pm 29.18 SD μ V, range =200-300 μ V). We then used the infomax ICA
555 algorithm after first using Principal Components Analysis (PCA) to reduce the number of channels from
556 an average of 67 to 32 for each subject to further speed ICA. ICA components constituting eye blinks,
557 eye movements, and heart artifacts (when present), were manually selected for rejection (mean=2.79
558 \pm 0.66 SD components, range=2-4 components). The resulting ICA weight matrix was then transferred
559 to the original dataset, after first rejecting the same bad channels in the original dataset as were
560 identified in the ICA dataset.

561 After the data was cleaned with ICA, we down sampled it to 512 Hz and applied a non-causal
562 0.5 to 40 Hz 8th order Butterworth bandpass filter using ERPLAB's filter function. Data were then re-
563 referenced to the average of the left and right earlobes. In the case that one or both earlobe channels
564 were rejected as bad, the average of the left and right mastoid electrodes was used as reference
565 instead. While re-referencing, activity at channel Cz (the previous reference) was recalculated and
566 placed back into the data. We then used spherical interpolation on any channels rejected.

567 EEG data were epoched -50 to +500 ms from the first chirp of each word in each condition, with
568 the prestimulus period used as baseline. Each condition had roughly the same number of events
569 (MonoMale = 2417, DualTargetMale = 2392, DualMaskerMale = 2392), except for one subject in which
570 data recording ended early due to computer error (MonoMale = 1249, DualTargetMale = 2000,
571 DualMaskerMale = 2392). Epochs with artifacts were cleaned with a 2-step process: first, a voltage
572 threshold of \pm 300 μ V was applied to remove any epochs with extreme artifacts. Second, any remaining
573 epochs with voltage activity outside of 6 SD for a single channel, or 2 SD across all channels were
574 rejected (MonoMale: mean = 5.01 \pm 0.83% SD rejected epochs, range = 3.07-6.85%; DualTargetMale:
575 mean = 4.84 \pm 0.86% SD rejected epochs, range = 3.75-7.02%; DualMaskerMale: mean = 5.12 \pm 1.1%
576 SD rejected epochs, range = 2.9-8.12%)

577

578 *N1 extraction*

579 N1 component activity was captured as the mean amplitude at electrode Fz between 110 and
580 160 ms post chirp onset. The electrode site Fz was chosen based on previous research using the
581 Cheech paradigm (Backer et al., 2019) and literature on the scalp regions where auditory evoked
582 responses generally reach peak amplitudes (Kappenman & Luck, 2011). Given the N1 component
583 lacked a clear peak (as is typical during continuous speech presentation) and was usually followed by
584 a sustained negativity (Figures 3, 4), a mean amplitude over the 50 ms interval was used instead of
585 more traditional peak fitting approaches, with the time course chosen based on visual inspection of
586 mono and dual target conditions (Figure 4). Difference scores were calculated by subtracting mean dual
587 target N1 amplitudes from mean dual masker N1 amplitudes.

588

589 **Statistical Analysis**

590 All statistical analyses were performed using R version 2023.09.1+494 (Posit Software, PBC)
591 on macOS Sonoma 14.6.1. A paired t-test was performed to evaluate the statistical significance of
592 differences between target and masker N1 amplitudes. A linear multiple regression analysis was
593 conducted to examine how N1 amplitude differences between attended (target) and unattended
594 (masker) speech relates to cognitive and non-audiometric psychoacoustic predictors. The criterion
595 variable was the N1 amplitude difference in the continuous multi-talker spatial attention task. The five
596 predictors included: (1) median reaction time differences between the incongruent and congruent
597 conditions of the Flanker task, (2) median reaction time differences between the incongruent and
598 congruent conditions of the Stroop task, (3) median reaction time differences between TMT-B and TMT-
599 A, (4) letter recall accuracy from 3 and 4 letter trials of the Reading Span test, and (5) average threshold
600 of the Temporal Fine Structure task across 250 and 500 Hz conditions.

601 The five predictors were entered simultaneously into the multiple regression model to assess
602 their collective effect on the N1 amplitude differences. Prior to the analysis, several diagnostic checks
603 confirmed the suitability of multiple regression. Residual vs. fitted plots and Q-Q plots indicated no
604 significant deviations from normality. Multicollinearity among predictors was assessed using variance
605 inflation factors (VIFs), with values above 5 suggesting problematic multicollinearity (Marcoulides &
606 Raykov, 2019). The highest VIF in our model was 1.43, indicating no substantial effects of
607 multicollinearity. Influential outliers were evaluated using a Cook's Distance threshold of 0.5. As
608 mentioned previously, one influential outlier was detected and subsequently removed, resulting in a
609 final sample size of $N=24$. Plots of each predictor against N1 amplitude differences demonstrated no
610 significant deviations from linearity. With a sample size of 24 and five predictors, the subject-per-variable
611 ratio (SPV) was 4.8. A linear regression model requires a minimum SPV of 2 for reliable estimation of
612 regression coefficients, standard errors, and confidence intervals (Austin & Steyerberg, 2015). All
613 variables were z-standardized prior to regression analysis.

614 Following the multiple regression analyses, dominance analysis was performed to assess the
615 relative importance of predictors in explaining the variance in N1 amplitude differences (Budescu,
616 1993). This method evaluates the dominance of each predictor by comparing their R^2 contributions
617 across all possible subset models. Specifically, each predictor is added to all possible models that do
618 not already include it, and the increase in R^2 is measured. This additional analysis was necessary
619 because the regression coefficients can be influenced by even moderate collinearity among predictors,
620 potentially skewing the assessment of each predictor's contribution (Mizumoto, 2023). A predictor's
621 general dominance weight (GDW; Azen and Budescu, 2003) is a measure of the proportion of the
622 variance explained by each predictor. It is calculated by averaging the squared semipartial correlations
623 for the predictor across all possible subset models. Higher GDW values indicate greater explanatory
624 power of a predictor when combined with other predictors. The total sum of the GDWs equals the
625 proportion of variance explained by the regression model, R^2 . Similar analysis approaches have been
626 used in previous studies, such as Oberfeld and Klöckner (2016) and Oberfeld et al. (2024).

627 To validate the predictive accuracy of the regression model, we employed Leave-One-Out Cross
628 Validation (LOOCV). This technique was utilized to ensure the model's robustness in predicting N1

629 amplitude differences based on participants' cognitive and psychophysical performance. In LOOCV,
630 one observation in the dataset (one participant) is used as a validation set, while the remaining datasets
631 (N-1) are used as training data to predict the value for the omitted observation. This iterative process is
632 repeated N times, once for each observation. For each iteration, the model's prediction error is
633 calculated, and these errors are aggregated to compute the root-mean-squared error (RMSE), which
634 provides a measure of the average magnitude of prediction errors. This method provides a nearly
635 unbiased estimate of the model's predictive performance, particularly beneficial for smaller datasets.

636 A separate multiple regression analysis was conducted using the same five predictors to
637 examine their ability to predict the proportion of color word hits in the continuous multi-talker spatial
638 attention task. The predictors were entered simultaneously into the regression model to assess their
639 combined impact on speech perception performance in noise. Prior to the analysis, the appropriateness
640 of using multiple regression was verified through several diagnostic checks. Although the residuals vs.
641 fitted plots and Q-Q plots indicated some deviations from normality, no transformation of the response
642 variable significantly improved the distribution of the residuals or altered the results. Since the same
643 predictors were used, the multicollinearity checks from the first model remained unchanged. The
644 influential outlier observed in the first model was also removed in this analysis (N=24). Plots of each
645 predictor against N1 amplitude differences showed no significant deviations from linearity. All variables
646 were z-standardized prior to regression analysis.

647

648

RESULTS

649

650 Cognitive and Psychoacoustic Predictors

651 Figure 2 presents box plots illustrating the distribution of scores across the five cognitive and
652 psychoacoustic tasks administered to participants, highlighting the variability in performance. While
653 some plots reveal visually apparent outliers, these data points did not influence the predictive
654 performance of the models and were therefore retained in the analysis.

655

656

[FIGURE 2]

657

658 N1 Across Conditions

659 The grand average waveforms in Figure 3A reveal a significant difference in N1 amplitudes
660 between the target and masker story conditions. The N1 amplitude was significantly larger for the target
661 story condition ($M = -0.142 \mu\text{V}$, $SD = 0.350 \mu\text{V}$) compared to the masker story condition ($M = 0.113 \mu\text{V}$,
662 $SD = 0.198 \mu\text{V}$). A paired t-test confirmed this difference as statistically significant ($t(23) = -3.664$, $p =$
663 0.001). Figure 3B displays the difference wave, created by subtracting the ERP waveform of the masker
664 from that of the target. Negative N1 amplitude differences indicate a stronger N1 response to target
665 speech compared to masker speech, while positive differences suggest a stronger response to the

666 masker than to the target. The average N1 amplitude difference between the target and masker speech
 667 was $-0.254 \pm 0.340 \mu\text{V}$, indicating a stronger (more negative) N1 response to the target speech
 668 compared to the masker.

669

670

[FIGURE 3]

671

672 After analyzing the grand average waveforms for the target and masker story conditions, we
 673 compared them to a baseline mono condition, in which there was only one speaker for participants to
 674 attend to. This comparison was conducted to determine whether selective attention in a dual-talker
 675 scenario primarily boosts the neural response to the target speech, suppresses the response to
 676 competing speech, or involves both processes. The N1 amplitude did not differ significantly between
 677 the (dual) target and mono story conditions, as indicated by a Bonferroni-corrected paired t-test ($t(23)$
 678 $= -0.790$, $p = 0.438$; see Figures 4A and 4B). However, the N1 amplitude for the masker story condition
 679 was significantly smaller than that of the mono story condition ($t(23) = -2.798$, $p = 0.010$), as shown in
 680 Figures 4C and 4D.

681

682

[FIGURE 4]

683

684 **Predicting N1 Amplitude Differences**

685 We used multiple regression analysis to assess whether performance on cognitive and
 686 psychoacoustic tasks could predict N1 amplitude differences between target and masker narratives
 687 (i.e., dual-talker condition only). As shown in Table 1, the multiple regression model demonstrated a
 688 strong fit.

689

690

[TABLE 1]

691

692

693 The N1 amplitude difference between target and masker narratives was significantly positively
 694 associated with Reading Span Accuracy on three and four-letter trials. This indicates that better working
 695 memory performance was associated with reduced differentiation in auditory cortical responses (i.e.,
 696 closer to 0 or more positive N1 amplitude difference between target and masker). In contrast, the N1
 697 amplitude difference demonstrated the opposite relationship with Stroop Task reaction time differences
 698 between incongruent and congruent trials. Specifically, stronger behavioral inhibition abilities, in the
 699 form of reduced response interference (i.e., *smaller* Stroop RT difference values), were associated with
 700 more robust attentional effects in the auditory cortex (i.e., larger target- compared to masker-evoked
 701 N1 responses).

702

703 TFS threshold and reaction time differences between incongruent and congruent trials on the
 Flanker and Trail Making Test were not statistically significant predictors. However, there was a trend

704 suggesting that better TFS ability (indicated by lower thresholds) was associated with smaller N1
705 amplitude differences. Figure 5 illustrates the relationships between each individual predictor and N1
706 difference amplitudes. Additionally, Figure 6 provides further insights by depicting the confidence
707 intervals for each predictor's coefficients.

708

709

[FIGURE 5]

710

711

[FIGURE 6]

712

713 Leave-One-Out Cross Validation (LOOCV) was used to validate the model fit and assess its
714 predictive performance. The scatter plot in Figure 7 illustrates the relationship between the actual N1
715 amplitude differences (x-axis) and the predicted N1 amplitude differences (y-axis) generated from the
716 LOOCV procedure. Each point on the plot represents an individual observation used as the validation
717 set in a single iteration of LOOCV. The diagonal line indicates perfect prediction where predicted values
718 match the actual values. Points closely aligned with this line indicate accurate predictions, while points
719 further away indicate prediction errors. The LOOCV analysis resulted in a root-mean-square error
720 (RMSE) of 0.215 microvolts, reflecting the average magnitude of prediction errors for the N1 amplitude
721 differences.

722

723

[FIGURE 7]

724

725 As previously mentioned, none of the six predictors exhibited strong collinearity, as indicated by
726 low Variance Inflation Factors (VIFs), suggesting that the model and its parameter variances should be
727 stable. Nonetheless, the predictors were mildly correlated (see Table 2). In such situations, relying
728 solely on regression coefficients to assess predictor importance can be misleading. To address this, we
729 employed the "dominance analysis" method (Budescu, 1993) to evaluate the relative importance of
730 each predictor in the regression model. The results revealed that Reading Span had the highest general
731 dominance weight (GDW), followed by Stroop, indicating that working memory and response inhibition
732 play significant roles in driving neural encoding differences between target and masker speech (see
733 Table 1). Notably, these findings align with the results of the multiple regression analysis, where
734 Reading Span emerged as the most significant predictor, followed by Stroop. Additionally, TFS showed
735 a relatively high GDW, consistent with its marginal significance ($p = 0.07$) in the regression model.

736

737

[TABLE 2]

738

739

740

741

742

743

744

745

Reliability measures were computed for each predictor variable. Cronbach's alpha values were 0.88 for the Flanker task, 0.67 for the Stroop task, and 0.75 for the Trail Making task. For the Reading Span task, KR20 was 0.67. Due to the nature of the adaptive procedure, TFS thresholds were not assessed for reliability. The reliability of the N1 amplitude difference was assessed using the intraclass correlation coefficient (ICC = 0.353), calculated based on two target and two masker stories per participant.

746

747

748

749

Finally, we also conducted bivariate correlations to examine the relationships between each predictor and N1 amplitude differences. The correlation plots summarizing these relationships are presented in Figure 8.

750

[FIGURE 8]

751

752

753

Predicting Behavioral Performance of Speech Perception in Noise

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

The earlier models explored the relationships between neural responses and performance metrics obtained separately from the EEG session. However, during the EEG session, participants also engaged in behavioral tasks that approximate real-world listening abilities. To further explore these relationships, we conducted an additional analysis using the same predictors to predict the proportion of color word hits, defined as the proportion of correctly identified color words in the target story within two seconds of their onset (i.e., excluding overlapping color words across the story pairs; Figure 9). This measure was selected because it directly reflects the participants' ability to focus on the target speech and respond promptly, making it a relevant and sensitive measure of real-world listening performance. We aimed to assess how well the observed relationships of the predictors with neural activity translate to behavioral outcomes. Based on this multiple regression model, none of the predictors were correlated with the proportion of color word hits (Table 3). This finding suggests that while cognitive and psychoacoustic abilities can accurately predict neural activity during SiN perception, they do not appear to predict SiN behavioral performance in this cohort. The relationship between each predictor and the proportion of color word hits is illustrated in Figure 10.

769

[FIGURE 9]

770

771

772

[TABLE 3]

773

774

775

[FIGURE 10]

776

777

Relationship Between N1 Amplitude Differences and Proportion of Color Word Hits

778

779

To gain further insight into the relationship between neural processing and speech perception, we performed a simple linear regression between the N1 amplitude differences and the proportion of

780 color words that participants correctly identified. N1 amplitude differences between target and masker
781 speech were not significantly associated with this ecologically valid SiN performance metric (Figure 11).

782

783

[FIGURE 11]

784

785

DISCUSSION

786

787 In this study, we aimed to identify the key predictors of neural processing within the auditory
788 cortex (specifically N1 amplitudes) associated with differentiating target from masker speech in a multi-
789 talker spatial attention task among normal hearing listeners. For clarity, we adhered to well-established
790 interpretations of each cognitive predictor, emphasizing the constructs most relevant to our dynamic
791 speech-in-noise (SiN) EEG task. Specifically, Reading Span gauged working memory abilities, the
792 Stroop task measured linguistic interference and inhibitory control, the Flanker task assessed selective
793 attention and stimulus inhibition, and the Trail Making Test evaluated broader executive functions,
794 including cognitive-attentional switching and processing speed. Additionally, Temporal Fine Structure
795 (TFS), a psychoacoustic measure, was included as it reflects sound localization abilities.

796

Predicting N1 Amplitudes

797 Our multiple regression model, which included five cognitive and non-audiometric
798 psychophysical predictors, demonstrated that working memory performance and inhibitory control
799 significantly contributed to the variance in N1 amplitude differences between target and masker talkers.
800 TFS also showed a trend toward further explaining additional variance. These findings demonstrate
801 that, even among listeners with normal hearing, neural encoding of auditory selective attention during
802 SiN listening varies based on domain-general abilities. This highlights the complexity of neural
803 mechanisms involved in auditory selective attention during SiN, which span from basic auditory
804 processing to higher-level cognitive functions.

805 Our multivariate approach further examines the relationship between each predictor and N1
806 amplitude differences between target and masker stories. In our study, better performance on the
807 Stroop task—indicated by smaller reaction time differences between incongruent and congruent
808 conditions—was associated with larger N1 amplitude differences between target and masker stories.
809 This finding aligns with our hypothesis that better cognitive abilities, such as response inhibition, lead
810 to larger N1 amplitude differences due to more effective suppression of distractor information and
811 improved attention to target speech. The Stroop task is designed to assess linguistic interference and
812 inhibitory control (Stroop, 1935; Troyer et al., 2006). This task requires participants to inhibit the
813 automatic tendency to read the word and instead name the ink color, demonstrating their ability to
814 suppress irrelevant information. This skill is likely transferable to complex listening environments, where
815 participants must focus on target speech while inhibiting masker speech. Individuals with stronger
816 inhibitory control may be better at suppressing neural representations of distracting information,
817 allowing for enhanced focus on the target speech.

818

819 To test whether the N1 amplitude difference was driven by target enhancement or masker
820 suppression, we compared the average N1 amplitudes of the dual-talker conditions to the single-talker
821 (mono) condition within the continuous multi-talker spatial attention task. The average N1 amplitude for
822 the masker stream was smaller than the single-talker condition. In the single-talker condition, there is
823 no need for inhibitory control as there are no competing stimuli. However, in a dual-talker scenario,
824 effective inhibition of irrelevant speech is crucial. Thus, these results indicate individuals with better
825 inhibitory control have significantly inhibited N1 responses to masker streams. This supports previous
826 findings showing that the N1 is sensitive to attentional processes (Luck, 1995), with larger N1 amplitude
827 differences reflecting greater attentional resource allocation for the target speech (Kerlin et al., 2010;
828 Picton & Hillyard, 1974; Stapells, 2009).

829 Notably, Accuracy on the Reading Span task for three- and four-letter trials was a significant
830 predictor, emphasizing the role of working memory in neural encoding processes during challenging
831 listening scenarios. Interestingly, higher accuracy on the Reading Span task was associated with
832 *smaller* N1 amplitude differences, contrary to our hypothesis. Individuals with stronger working memory
833 abilities are likely better at holding and manipulating information about the target speech. As a result,
834 they may be more capable of using this information to predict upcoming auditory input, which is crucial
835 for understanding speech, particularly in noisy environments (Akeroyd, 2008; Obleser & Eisner, 2009).
836 These predictive processes might reduce the need for effortful processing of target speech, as the brain
837 can rely on expectations for comprehension. Conversely, individuals with poorer working memory may
838 struggle to predict and pre-process information in the target stream, therefore requiring a greater
839 reliance on real-time processing. This may increase cognitive load and demand more neural resources
840 for target speech focus and distractor suppression, potentially explaining larger N1 amplitude
841 differences in these individuals.

842 Interestingly, in our pilot testing, the total Reading Span accuracy of all trials (3 to 7 sentence-
843 letter pairs) was not a strong predictor of N1 amplitude differences between target and masker speech.
844 This finding suggests that accuracy on three- and four-letter trials may be a more informative measure
845 of working memory capacity related to neural processes in SiN perception. One possible explanation is
846 that tasks involving fewer items (three to four letters) fall within the working memory capacity limit
847 (Cowan, 2001, 2015). This smaller capacity measure might be more sensitive to individual differences
848 in neural efficiency and attentional control during auditory attention tasks. When the Reading Span
849 exceeds this capacity (five to seven letters), performance may be influenced by additional factors such
850 as strategies, chunking, or long-term memory processes, which may not directly reflect the working
851 memory capacity relevant to real-time speech processing in noise. Consequently, focusing on moderate
852 working memory capacities may provide a more accurate representation of the cognitive resources
853 used during complex listening tasks, leading to better predictions of neural activity differences in multi-
854 talker auditory environments.

855 Temporal Fine Structure (TFS) thresholds did not significantly predict N1 amplitude differences.
856 Nevertheless, as shown in Figure 5, there is a notable trend suggesting that better TFS thresholds are
857 associated with smaller N1 amplitude differences between target and masker speech. This trend
858 contradicts our initial hypothesis, where we expected better TFS sensitivity to correspond to larger N1
859 amplitude differences. TFS processing begins in the peripheral auditory system, starting with basilar
860 membrane filters in the cochlea and extends through the brainstem (Borjigin, 2023; Plomp, 1964).

861 These processes are important for encoding interaural timing differences, which are crucial for
862 separating target and masker speech based on azimuthal location. In SiN perception, this suggests that
863 TFS processing contributes to the early spatial segregation of auditory streams, likely occurring at the
864 cochlea and brainstem levels. Individuals with better TFS sensitivity may exhibit larger neural
865 differentiation of speech streams in these early stages of auditory processing. This enhanced peripheral
866 processing may reduce the need for differentiation or attentional modulation in the cortex, as the target
867 and masker streams may already be separated when they reach the auditory cortex. This may explain
868 why better TFS sensitivity is associated with smaller N1 amplitude differences, reflecting less cortical
869 effort for stream segregation. TFS was tested at a fundamental frequency (F0) that differs from the F0
870 of the male talkers used in this study. Specifically, the TFS F0 was selected to be approximately halfway
871 between the male and female (re-synthesized) voices identified during pilot recordings. While this
872 choice allowed us to test TFS performance in a frequency range relevant to both genders, it introduces
873 a potential limitation in directly relating TFS thresholds to the cortical processing of male voices. This
874 highlights the need for further research to better understand the relationship between peripheral
875 auditory processing and cortical neural responses, particularly in challenging listening scenarios.

876 The Trail Making Test (TMT) reaction times were not significant predictors of N1 amplitude
877 differences in our study. The TMT primarily assesses executive functioning in a visuo-motor domain by
878 requiring participants to switch attention between numerical and alphabetical sequences (e.g.,
879 connecting 1 to A, 2 to B) in ascending order. While our multi-talker spatial attention task also involves
880 attentional switching, the type of switching required in the TMT may not align with the demands of
881 auditory attention in speech-in-noise perception. The attentional switching in our continuous multi-talker
882 task involves dynamically redirecting attention between left and right ears as the target story alternates
883 between spatial locations throughout its narration. This process requires not only attentional control but
884 also the integration of spatial auditory cues to track the target stream while suppressing competing
885 distractor speech. These demands are likely supported by domain-specific auditory and linguistic
886 mechanisms, which involve distinct neural pathways optimized for processing spatial and linguistic
887 auditory information. This distinction highlights the importance of task-specific predictors when studying
888 complex auditory tasks. Further research could explore whether attention-switching measures within
889 the auditory domain, or involving linguistic components, provide stronger associations with neural
890 indices like N1 amplitude differences.

891 Flanker test reaction times were not significant predictors of N1 amplitude differences in our
892 study. Although the Flanker test is a well-established measure of selective attention in (visual) space, it
893 may not fully capture the neural processes required to differentiate between target and masker speech
894 in a noisy environment. The Flanker, like the Stroop, also assesses inhibitory control, but does so in a
895 different context. The Stroop test requires participants to inhibit a more automatic linguistic response
896 (reading the word) in favor of a less automatic one (naming the ink color), directly engaging cognitive
897 mechanisms related to speech and language processing. This linguistic interference is evidently
898 relevant to the challenges of SiN perception, especially when maskers are meaningful speech rather
899 than non-linguistic sounds like steady-state noise, which does not trigger lexical activity (Lu et al., 2016;
900 Schneider et al., 2022). In contrast, the Flanker test primarily involves visuo-spatial interference, where
901 participants must inhibit responses to maskers visually adjacent to the target stimulus. This suggests
902 that in individuals with normal hearing performing a realistic SiN task, linguistic interference is more

903 impactful than (visuo-)spatial interference. Future work might determine whether this asymmetry also
904 holds when listeners have hearing loss and, consequently, greater challenges with spatial masking.

905

906 **Implications for the N1**

907 Our findings suggest a complex interaction between cognitive abilities and auditory processing,
908 updating our understanding of the N1 component. Traditionally considered a marker of basic auditory
909 processing and attention, the N1 also appears to be influenced by higher-order cognitive functions such
910 as inhibition and working memory, particularly in challenging listening scenarios. Extensive research
911 suggests that when individuals attend to specific auditory stimuli, the N1 amplitude is typically
912 enhanced, indicating a greater neural response to the attended sounds (Hillyard et al., 1973; Kerlin et
913 al., 2010; Picton & Hillyard, 1974; Stapells, 2009). Consistent with this, better response inhibition, as
914 measured by the Stroop task, is associated with larger N1 amplitude differences. However, our findings
915 from the Reading Span test indicate that this relationship is more complicated; higher working memory
916 capacity is associated with *smaller* N1 amplitude differences between attended and unattended
917 streams. This suggests that neural processing during speech perception is determined by
918 compensations and compromises between top-down cognitive influences.

919

920 **Predicting Behavioral Performance of Speech Perception in Noise**

921 In addition to analyzing the direct relationship between cognitive performance and neural
922 mechanisms of auditory selective attention (N1 amplitude differences), we also assessed how each
923 cognitive predictor relates to ecologically relevant listening behaviors during the continuous multi-talker
924 spatial attention task, conducted simultaneously while EEG was recorded. Specifically, we tested the
925 relationship of each cognitive predictor and the proportion of correctly identified color words in the target
926 story. Interestingly, none of the cognitive and psychoacoustic predictors showed a significant correlation
927 with this behavioral measure.

928 This lack of a significant correlation between cognitive abilities and SiN perception is consistent
929 with previous findings in normal-hearing individuals. Füllgrabe & Rosen, 2016 found that working
930 memory capacity is associated with SiN intelligibility in older, hearing-impaired individuals, but this
931 correlation is absent in younger, normal-hearing listeners. One possible explanation is that age-related
932 auditory deficits force older, hearing-impaired individuals to rely more heavily on working memory-based
933 compensatory mechanisms for successful SiN perception. In contrast, younger listeners may depend
934 less on cognitive resources due to better auditory processing. However, other studies, such as Oberfeld
935 & Klöckner-Nowotny, 2016, have demonstrated significant associations between cognition and SiN
936 perception. The meta-analysis by Dryden et al. (2017) highlights three key factors that could explain
937 these mixed findings: differences in cognitive measures, variability in SiN tasks, and participants'
938 hearing thresholds. Future research should consider these factors to better understand the nuanced
939 role of cognition in SiN processing.

940 Additionally, N1 amplitude differences were not associated with SiN perception (proportion of
941 color word hits) in our study. The N1 component has been widely associated with SiN perception in
942 previous studies (Bidelman & Howell, 2016; Billings et al., 2013; Parbery-Clark et al., 2011). The lack
943 of association between N1 amplitudes and behavioral speech-in-noise (SiN) performance in our study
944 may be due to differences in the tasks used to assess SiN perception. Conventional measures such as

945 the Hearing in Noise Test (Parbery-Clark et al., 2011), QuickSiN (Bidelman & Howell, 2016), and
946 sentence-in-noise identification tasks (Billings et al., 2013) assess word or sentence identification in
947 noise. These tasks require participants to process and repeat semantically meaningful speech. In
948 contrast, our SiN task required participants to detect and respond to embedded color words in a
949 continuous multi-talker spatial attention paradigm, measuring their ability to selectively attend to the
950 target stream and suppress distractors. This task emphasized rapid attentional shifts and target word
951 detection, rather than the linguistic reconstruction and sentence-level recognition typically evaluated in
952 more conventional SiN measures. Task differences may explain the lack of correlation between N1
953 amplitudes and SiN perception, emphasizing the need to consider experimental design when
954 interpreting neural-behavioral relationships. Future studies should investigate how task demands and
955 behavioral measures assess distinct facets of SiN processing.

956 Although cognitive abilities and N1 amplitudes *were not* significantly associated with listening
957 performance, it is noteworthy that cognitive abilities *were* significantly associated with N1 amplitudes.
958 These findings highlight an important distinction between neural encoding of auditory selective attention
959 and behavioral SiN performance. N1 amplitude differences reflect underlying neural mechanisms
960 involved in focusing attention on target speech and suppressing masker speech, whereas behavioral
961 SiN performance measured in our task likely involves additional factors and compensatory strategies.
962 Given that SiN processing is multifaceted, no single behavioral speech test, such as color word
963 detection, can readily parse the many underlying mechanisms that might contribute to it. In these cases,
964 neural metrics can add substantially to our understanding which might not be fully captured by
965 behavioral data alone.

966 Additionally, our cohort consisted of young, highly educated individuals, with all participants
967 having completed at least 13 years of education. It is possible that the well-educated status of these
968 participants equips them with advanced compensatory mechanisms that may enable strong
969 performance in speech perception tasks, despite differences in cognitive abilities. As shown in Figure
970 8, participants achieved a high proportion of color word hits, indicating strong overall performance. For
971 those with lower cognitive task performance, their educational background and overall health may have
972 provided them with effective strategies for maintaining high performance in challenging listening tasks.
973 These strategies might include better use of contextual cues, more effective allocation of attentional
974 resources, or superior problem-solving skills, allowing them to compensate for cognitive limitations.
975 Consequently, the homogeneity in educational background might mask differences in cognitive
976 performance, resulting in consistently strong performance in the continuous multi-talker spatial attention
977 task, regardless of cognitive predictor scores. Although our data do not allow definitive claims, the
978 findings suggest that N1 amplitude differences might serve as early markers of auditory deficits or
979 hearing loss, warranting further investigation in diverse populations.

980

981 **Limitations**

982 A key limitation of this study is the relatively small sample size of 25 participants, which may
983 limit the ability to fully capture the complexity and variability in individual differences, especially with
984 multiple regression models involving five predictors. To improve the reliability and generalizability of
985 these findings, future studies should prioritize larger samples to better represent the population and
986 enhance the robustness of the results. Another limitation of this study is the scoring approach used for

987 the working memory span task. We acknowledge that this method differs from the traditional scoring
988 approach, which averages accuracy across all list lengths (Conway et al., 2005). Our choice was guided
989 by theoretical accounts of working memory capacity (Cowan, 2001, 2015) and the need for a stable
990 measure that reflected meaningful individual differences. Future research should explore how different
991 scoring approaches impact findings to ensure robustness across methodologies. Additionally, in this
992 study, the reliability of cognitive performance measures was adequate, with Cronbach's alpha and KR20
993 values ranging from 0.67 to 0.88. For the criterion measure, the N1 amplitude difference, the ICC was
994 0.353. This relatively low ICC may be partly attributable to the small sample size, which inherently
995 increase variability and reduce the precision of reliability estimates. Future studies should aim to
996 address this limitation by including a larger sample size to improve stability of reliability estimates.
997

998 **Conclusion**

999 Overall, our findings highlight the complexity of the relationships between auditory processing,
1000 cognitive abilities, and attentional gains in speech perception under noisy conditions. We demonstrate
1001 that neural measures reflect impactful inter-individual differences that are not fully captured by solely
1002 traditional assessments of audiometric hearing and behavioral performance. These findings have
1003 significant implications for improving hearing loss diagnostics, which currently rely primarily on ear
1004 health assessments. By incorporating cognitive abilities and neural activity measures, we can obtain a
1005 more comprehensive understanding of speech perception, even among normal hearing listeners.
1006 Having established the predictive value of these tasks for neural processing of SiN, future work might
1007 further clarify *how* each of these reflects differences in underlying cognitive mechanisms across
1008 individuals. Understanding how the brain processes auditory information in noisy environments—and
1009 how differences in cognition influence this process—could lead to earlier detection of hearing loss,
1010 improved strategies for enhancing speech perception in daily life, and ultimately, a significant
1011 improvement in the quality of life for individuals facing hearing challenges.

1012 **References**

1013

- 1014 1. Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual
 1015 differences in cognitive ability? A survey of twenty experimental studies with normal and
 1016 hearing-impaired adults. *International Journal of Audiology*, 47(sup2), S53–S71.
 1017 <https://doi.org/10.1080/14992020802301142>
- 1018 2. Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in*
 1019 *Bioscience: A Journal and Virtual Library*, 5, D202-212. <https://doi.org/10.2741/alain>
- 1020 3. Armstrong, C., Thresh, L., Murphy, D., & Kearney, G. (2018). A Perceptual Evaluation of
 1021 Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database. *Applied*
 1022 *Sciences*, 8(11), Article 11. <https://doi.org/10.3390/app8112029>
- 1023 4. Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in
 1024 linear regression analyses. *Journal of Clinical Epidemiology*, 68(6), 627–636.
 1025 <https://doi.org/10.1016/j.jclinepi.2014.12.014>
- 1026 5. Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing
 1027 predictors in multiple regression. *Psychological Methods*, 8(2), 129–148.
 1028 <https://doi.org/10.1037/1082-989X.8.2.129>
- 1029 6. Backer, K. C., Kessler, A. S., Lawyer, L. A., Corina, D. P., & Miller, L. M. (2019). A novel EEG
 1030 paradigm to simultaneously and rapidly assess the functioning of auditory and visual
 1031 pathways. *Journal of Neurophysiology*, 122(4), 1312–1329.
 1032 <https://doi.org/10.1152/jn.00868.2018>
- 1033 7. Bednar, A., & Lalor, E. C. (2020). Where is the cocktail party? Decoding locations of attended
 1034 and unattended moving sound sources using EEG. *NeuroImage*, 205, 116283.
 1035 <https://doi.org/10.1016/j.neuroimage.2019.116283>
- 1036 8. Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric
 1037 cortical processing underlying degraded speech perception. *NeuroImage*, 124(Pt A), 581–590.
 1038 <https://doi.org/10.1016/j.neuroimage.2015.09.020>
- 1039 9. Billings, C.J., McMillan, G.P., Penman, T.M. et al. Predicting Perception in Noise Using
 1040 Cortical Auditory Evoked Potentials. *JARO* 14, 891–903 (2013).
 1041 <https://doi.org/10.1007/s10162-013-0415-y>
- 1042 10. Borjigin, A. (2023). *The Role of Temporal Fine Structure in Everyday Hearing* [Thesis, Purdue
 1043 University Graduate School]. <https://doi.org/10.25394/PGS.19673883.v1>
- 1044 11. Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
 1045 <https://doi.org/10.1163/156856897X00357>
- 1046 12. Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).
 1047 Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of
 1048 Natural, Narrative Speech. *Current Biology: CB*, 28(5), 803-809.e3.
 1049 <https://doi.org/10.1016/j.cub.2018.01.080>
- 1050 13. Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative
 1051 importance of predictors in multiple regression. *Psychological Bulletin*, 114(3), 542–551.
 1052 <https://doi.org/10.1037/0033-2909.114.3.542>

- 1053 14. Campbell, T., Kerlin, J. R., Bishop, C. W., & Miller, L. M. (2012). Methods to Eliminate
 1054 Stimulus Transduction Artifact From Insert Earphones During Electroencephalography. *Ear*
 1055 *and Hearing*, 33(1), 144. <https://doi.org/10.1097/AUD.0b013e3182280353>
- 1056 15. Cooper, J. C., & Gates, G. A. (1991). Hearing in the Elderly-The Framingham Cohort, 1983-
 1057 1985: Part II. Prevalence of Central Auditory Processing Disorders. *Ear and hearing*, 12(5),
 1058 304–311. <https://doi.org/10.1037/h0043158>
- 1059 16. Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental
 1060 storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
 1061 <https://doi.org/10.1017/S0140525X01003922>
- 1062 17. Cowan, N. (2015). George Miller's Magical Number of Immediate Memory in Retrospect:
 1063 Observations on the Faltering Progression of Science. *Psychological Review*, 122(3), 536–
 1064 541. <https://doi.org/10.1037/a0039035>
- 1065 18. Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and
 1066 reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
 1067 [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- 1068 19. Davidson, A., & Souza, P. (2024). Relationships Between Auditory Processing and Cognitive
 1069 Abilities in Adults: A Systematic Review. *Journal of speech, language, and hearing research:*
 1070 *JSLHR*, 67(1), 296–345. https://doi.org/10.1044/2023_JSLHR-22-00716
- 1071 20. Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial
 1072 EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*,
 1073 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- 1074 21. Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while
 1075 listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29),
 1076 11854–11859. <https://doi.org/10.1073/pnas.1205381109>
- 1077 22. Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between
 1078 Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic
 1079 Literature Review and Meta-Analysis. *Trends in Hearing*, 21, 2331216517744675.
 1080 <https://doi.org/10.1177/2331216517744675>
- 1081 23. Füllgrabe, C., & Moore, B. C. J. (2018). The Association Between the Processing of Binaural
 1082 Temporal-Fine-Structure Information and Audiometric Threshold and Age: A Meta-Analysis.
 1083 *Trends in Hearing*, 22, 2331216518797259. <https://doi.org/10.1177/2331216518797259>
- 1084 24. Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2015). Age-group differences in speech
 1085 identification despite matched audiometrically normal hearing: Contributions from auditory
 1086 temporal processing and cognition. *Frontiers in Aging Neuroscience*, 6.
 1087 <https://www.frontiersin.org/articles/10.3389/fnagi.2014.00347>
- 1088 25. Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the trail making
 1089 test: What makes part B harder? *Journal of Clinical and Experimental Neuropsychology*, 17(4),
 1090 529–535. <https://doi.org/10.1080/01688639508405143>
- 1091 26. Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by
 1092 temporally-sensitive neural measures. *WIREs Cognitive Science*, 12(2), e1541.
 1093 <https://doi.org/10.1002/wcs.1541>

- 1094 27. Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech
 1095 recognition performance among young and elderly listeners. *Journal of Speech, Language,*
 1096 *and Hearing Research: JSLHR*, *40*(2), 423–431. <https://doi.org/10.1044/jslhr.4002.423>
- 1097 28. Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews*
 1098 *Neuroscience*, *5*(11), 887–892. <https://doi.org/10.1038/nrn1538>
- 1099 29. Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., & Melcher, J. R. (2007). Human
 1100 cortical activity during streaming without spectral cues suggests a general neural substrate for
 1101 auditory stream segregation. *The Journal of Neuroscience: The Official Journal of the Society*
 1102 *for Neuroscience*, *27*(48), 13074–13081. <https://doi.org/10.1523/JNEUROSCI.2299-07.2007>
- 1103 30. Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L.,
 1104 & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning
 1105 and attention studied with positron emission tomography. *Intelligence*, *12*(2), 199–217.
 1106 [https://doi.org/10.1016/0160-2896\(88\)90016-5](https://doi.org/10.1016/0160-2896(88)90016-5)
- 1107 31. Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective
 1108 Attention in the Human Brain. *Science*, *182*(4108), 177–180.
 1109 <https://doi.org/10.1126/science.182.4108.177>
- 1110 32. Hind, S. E., Haines-Bazrafshan, R., Benton, C. L., Brassington, W., Towle, B., & Moore, D. R.
 1111 (2011). Prevalence of clinical referrals having hearing thresholds within normal limits.
 1112 *International Journal of Audiology*, *50*(10), 708–716.
 1113 <https://doi.org/10.3109/14992027.2011.582049>
- 1114 33. Holder, J. T., Levin, L. M., & Gifford, R. H. (2018). Speech recognition in noise for adults with
 1115 normal hearing: Age-normative performance for AzBio, BKB-SIN, and QuickSIN. *Otology &*
 1116 *Neurotology: Official Publication of the American Otological Society, American Neurotology*
 1117 *Society [and] European Academy of Otology and Neurotology*, *39*(10), e972–e978.
 1118 <https://doi.org/10.1097/MAO.0000000000002003>
- 1119 34. Hopkins, K., & Moore, B. C. J. (2010). Development of a fast method for measuring sensitivity
 1120 to temporal fine structure information at low frequencies. *International Journal of Audiology*,
 1121 *49*(12), 940–946. <https://doi.org/10.3109/14992027.2010.512613>
- 1122 35. Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing
 1123 speech in older adults. *Aging, Neuropsychology, and Cognition*, *19*(6), 741–758.
 1124 <https://doi.org/10.1080/13825585.2011.652590>
- 1125 36. Kappenman, E. S., & Luck, S. J. (2011). ERP Components: The Ups and Downs of Brainwave
 1126 Recordings. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related*
 1127 *Potential Components* (p. 0). Oxford University Press.
 1128 <https://doi.org/10.1093/oxfordhb/9780195374148.013.0014>
- 1129 37. Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech
 1130 representations using a pitch-adaptive time–frequency smoothing and an instantaneous-
 1131 frequency-based F0 extraction: Possible role of a repetitive structure in sounds1. *Speech*
 1132 *Communication*, *27*(3), 187–207. [https://doi.org/10.1016/S0167-6393\(98\)00085-5](https://doi.org/10.1016/S0167-6393(98)00085-5)
- 1133 38. Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech
 1134 analysis, modification and synthesis framework. *Sadhana*, *36*(5), 713–727.
 1135 <https://doi.org/10.1007/s12046-011-0043-3>

- 1136 39. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-
 1137 STRAIGHT: A temporally stable power spectral representation for periodic signals and
 1138 applications to interference-free spectrum, F0, and aperiodicity estimation. *2008 IEEE*
 1139 *International Conference on Acoustics, Speech and Signal Processing*, 3933–3936.
 1140 <https://doi.org/10.1109/ICASSP.2008.4518514>
- 1141 40. Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical
 1142 Speech Representations in a “Cocktail Party.” *Journal of Neuroscience*, *30*(2), 620–628.
 1143 <https://doi.org/10.1523/JNEUROSCI.3631-09.2010>
- 1144 41. Klein, M., Ponds, R. W. H. M., Houx, P. J., & Jolles, J. (1997). Effect of test duration on age-
 1145 related differences in stroop interference. *Journal of Clinical and Experimental*
 1146 *Neuropsychology*, *19*(1), 77–82. <https://doi.org/10.1080/01688639708403838>
- 1147 42. Lu, Z., Daneman, M., & Schneider, B. A. (2016). Does increasing the intelligibility of a
 1148 competing sound source interfere more with speech comprehension in older adults than it
 1149 does in younger adults? *Attention, Perception, & Psychophysics*, *78*(8), 2655–2677.
 1150 <https://doi.org/10.3758/s13414-016-1193-5>
- 1151 43. Luck, S. J. (1995). Multiple mechanisms of visual-spatial attention: Recent evidence from
 1152 human electrophysiology. *Behavioural Brain Research*, *71*(1), 113–123.
 1153 [https://doi.org/10.1016/0166-4328\(95\)00041-0](https://doi.org/10.1016/0166-4328(95)00041-0)
- 1154 44. Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT
 1155 Press.
- 1156 45. Luck, S. J. (2022). *Applied Event-Related Potential Data Analysis*. LibreTexts.
 1157 <https://doi.org/10.18115/D5QG92>
- 1158 46. Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of*
 1159 *Audiology*, *42 Suppl 1*, S49-58. <https://doi.org/10.3109/14992020309074624>
- 1160 47. Marcoulides, K. M., & Raykov, T. (2019). Evaluation of Variance Inflation Factors in
 1161 Regression Models Using Latent Variable Modeling Methods. *Educational and Psychological*
 1162 *Measurement*, *79*(5), 874–882. <https://doi.org/10.1177/0013164418817803>
- 1163 48. Martin, R. C. (2021). The Critical Role of Semantic Working Memory in Language
 1164 Comprehension and Production. *Current Directions in Psychological Science*, *30*(4), 283–291.
 1165 <https://doi.org/10.1177/0963721421995178>
- 1166 49. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our
 1167 capacity for processing information. *Psychological Review*, *63*(2), 81–97.
 1168 <https://doi.org/10.1037/h0043158>
- 1169 50. Miller, L. M., Moore, B. D., & Bishop, C. W. (2020). *Frequency-multiplexed speech-sound*
 1170 *stimuli for hierarchical neural characterization of speech processing* (United States Patent
 1171 PCT/US15/40629). <https://patents.google.com/patent/WO2016011189A1/en>
- 1172 51. Mizumoto, A. (2023). Calculating the Relative Importance of Multiple Regression Predictor
 1173 Variables Using Dominance Analysis and Random Forests. *Language Learning*, *73*(1), 161–
 1174 196. <https://doi.org/10.1111/lang.12518>
- 1175 52. Moore, B. C. J. (2008). The Role of Temporal Fine Structure Processing in Pitch Perception,
 1176 Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People. *JARO*:

- 1177 *Journal of the Association for Research in Otolaryngology*, 9(4), 399–406.
 1178 <https://doi.org/10.1007/s10162-008-0143-x>
- 1179 53. Moore, B. C. J. (2021). Effects of hearing loss and age on the binaural processing of temporal
 1180 envelope and temporal fine structure information. *Hearing Research*, 402, 107991.
 1181 <https://doi.org/10.1016/j.heares.2020.107991>
- 1182 54. Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL)
 1183 and PEBL Test Battery. *Journal of Neuroscience Methods*, 222, 250–259.
 1184 <https://doi.org/10.1016/j.jneumeth.2013.10.024>
- 1185 55. Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I.,
 1186 Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief
 1187 Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*,
 1188 53(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- 1189 56. Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention
 1190 predict speech identification at a cocktail party. *eLife*, 5, e16747.
 1191 <https://doi.org/10.7554/eLife.16747>
- 1192 57. Oberfeld, D., Staab, K., Kattner, F., & Ellermeier, W. (2024). Is Recognition of Speech in Noise
 1193 Related to Memory Disruption Caused by Irrelevant Sound?. *Trends in hearing*, 28,
 1194 23312165241262517. <https://doi.org/10.1177/23312165241262517>
- 1195 58. Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory
 1196 cortex. *Trends in cognitive sciences*, 13(1), 14–19. <https://doi.org/10.1016/j.tics.2008.09.005>
- 1197 59. Oh, Y., Bridges, S. E., Schoenfeld, H., Layne, A. O., & Eddins, D. (2021). Interaction between
 1198 voice-gender difference and spatial separation in release from masking in multi-talker listening
 1199 environments. *JASA Express Letters*, 1(8), 084404. <https://doi.org/10.1121/10.0005831>
- 1200 60. Oh, Y., Hartling, C. L., Srinivasan, N. K., Diedesch, A. C., Gallun, F. J., & Reiss, L. A. J.
 1201 (2022). Factors underlying masking release by voice-gender differences and spatial separation
 1202 cues in multi-talker listening environments in listeners with and without hearing loss. *Frontiers*
 1203 *in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.1059639>
- 1204 61. Parbery-Clark, A., Marmel, F., Bair, J. and Kraus, N. (2011), What subcortical–cortical
 1205 relationships tell us about processing speech in noise. *European Journal of Neuroscience*, 33:
 1206 549-557. <https://doi.org/10.1111/j.1460-9568.2010.07546.x>
- 1207 62. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming
 1208 numbers into movies. *Spatial Vision*, 10(4), 437–442.
 1209 <https://doi.org/10.1163/156856897X00366>
- 1210 63. Perrone-Bertolotti, M., Tassin, M., & Meunier, F. (2017). Speech-in-speech perception and
 1211 executive function involvement. *PLOS ONE*, 12(7), e0180084.
 1212 <https://doi.org/10.1371/journal.pone.0180084>
- 1213 64. R. Plomp; The Ear as a Frequency Analyzer. *J. Acoust. Soc. Am.* 1 September 1964; 36 (9):
 1214 1628–1636. <https://doi.org/10.1121/1.1919256>
- 1215 65. Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults
 1216 listen to and remember speech in noise. *The Journal of the Acoustical Society of America*,
 1217 97(1), 593–608. <https://doi.org/10.1121/1.412282>

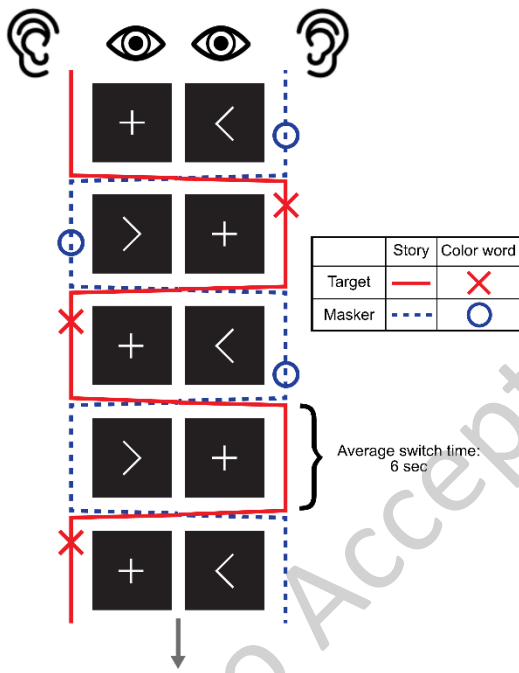
- 1218 66. Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of
 1219 attention. *Electroencephalography and Clinical Neurophysiology*, 36, 191–200.
 1220 [https://doi.org/10.1016/0013-4694\(74\)90156-4](https://doi.org/10.1016/0013-4694(74)90156-4)
- 1221 67. Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain
 1222 Damage. *Perceptual and Motor Skills*, 8(3), 271–276.
 1223 <https://doi.org/10.2466/pms.1958.8.3.271>
- 1224 68. Ronnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory
 1225 system for ease of language understanding (ELU). *Int J Audiol*, 47 Suppl 2, S99-105.
 1226 <https://doi.org/10.1080/14992020802301167>
- 1227 69. Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnerberg, J. (2012). Working Memory
 1228 Capacity May Influence Perceived Effort during Aided Speech Recognition in Noise. *Journal of*
 1229 *the American Academy of Audiology*, 23(08), 577–589. <https://doi.org/10.3766/jaaa.23.7.7>
- 1230 70. Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial Selective Auditory Attention in the
 1231 Presence of Reverberant Energy: Individual Differences in Normal-Hearing Listeners. *Journal*
 1232 *of the Association for Research in Otolaryngology*, 12(3), 395–405.
 1233 <https://doi.org/10.1007/s10162-010-0254-z>
- 1234 71. Schneider, B. A., Rabaglia, C., Avivi-Reich, M., Krieger, D., Arnott, S. R., & Alain, C. (2022).
 1235 Age-Related Differences in Early Cortical Representations of Target Speech Masked by Either
 1236 Steady-State Noise or Competing Speech. *Frontiers in Psychology*, 13.
 1237 <https://doi.org/10.3389/fpsyg.2022.935475>
- 1238 72. Şek, A. P., & Moore, B. C. J. (2012). Implementation of two tests for measuring sensitivity to
 1239 temporal fine structure. *International Journal of Audiology*, 51(1), 58–63.
 1240 <https://doi.org/10.3109/14992027.2011.605808>
- 1241 73. Şek, A. P., & Moore, B. C. J. (2021). *Guide to PSYCHOACOUSTICS*. Adam Mickiewicz
 1242 University Press.
- 1243 74. Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in*
 1244 *Cognitive Sciences*, 12(5), 182–186. <https://doi.org/10.1016/j.tics.2008.02.003>
- 1245 75. Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates
 1246 of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18(1), 1–13.
 1247 <https://doi.org/10.1162/089892906775250021>
- 1248 76. Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition
 1249 in young and older adults: The interaction of lexical competition and semantic context.
 1250 *Psychology and Aging*, 14(3), 458–472. <https://doi.org/10.1037//0882-7974.14.3.458>
- 1251 77. Stapells, D. (2009). *Cortical Event-Related Potentials to Auditory Stimuli*. (pp. 395–430).
- 1252 78. Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). Stroop Test. In *A Compendium of*
 1253 *Neuropsychological Tests: Administration, Norms, and Commentary* (3rd ed., pp. 477–499).
 1254 Oxford University Press.
- 1255 79. Strelcyk, O., Zahorik, P., Shehorn, J., Patro, C., & Derleth, R. P. (2019). Sensitivity to
 1256 Interaural Phase in Older Hearing-Impaired Listeners Correlates With Nonauditory Trail
 1257 Making Scores and With a Spatial Auditory Task of Unrelated Peripheral Origin. *Trends in*
 1258 *Hearing*, 23, 2331216519864499. <https://doi.org/10.1177/2331216519864499>

- 1259 80. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*
1260 *Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- 1261 81. Teoh, E. S., Ahmed, F., & Lalor, E. C. (2022). Attention Differentially Affects Acoustic and
1262 Phonetic Feature Encoding in a Multispeaker Environment. *Journal of Neuroscience*, 42(4),
1263 682–691. <https://doi.org/10.1523/JNEUROSCI.1455-20.2021>
- 1264 82. Teoh, E. S., & Lalor, E. C. (2019). EEG decoding of the target speaker in a cocktail party
1265 scenario: Considerations regarding dynamic switching of talker location. *Journal of Neural*
1266 *Engineering*, 16(3), 036017. <https://doi.org/10.1088/1741-2552/ab0cf1>
- 1267 83. The MathWorks Inc. (2021) MATLAB Version: 9.11.0.1769968 (R2021b), Natick,
1268 Massachusetts: The MathWorks Inc. <https://www.mathworks.com>
- 1269 84. Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., &
1270 Cruickshanks, K. J. (2015). Self-Reported Hearing Difficulties Among Adults With Normal
1271 Audiograms: The Beaver Dam Offspring Study. *Ear and Hearing*, 36(6), e290–e299.
1272 <https://doi.org/10.1097/AUD.0000000000000195>
- 1273 85. Troyer, A. K., Leach, L., & Strauss, E. (2006). Aging and Response Inhibition: Normative Data
1274 for the Victoria Stroop Test. *Aging, Neuropsychology, and Cognition*, 13(1), 20–35.
1275 <https://doi.org/10.1080/138255890968187>
- 1276 86. Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex
1277 working memory span tasks and higher-order cognition: A latent-variable analysis of the
1278 relationship between processing and storage. *Memory*, 17(6), 635–654.
1279 <https://doi.org/10.1080/09658210902998047>
- 1280 87. Winkler, I., Denham, S., & Escera, C. (2013). Auditory Event-related Potentials. In D. Jaeger &
1281 R. Jung (Eds.), *Encyclopedia of Computational Neuroscience* (pp. 1–29). Springer.
1282 https://doi.org/10.1007/978-1-4614-7320-6_99-1
- 1283 88. Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnerberg, J. (2013). The effects of working
1284 memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the*
1285 *Acoustical Society of America*, 134(3), 2225–2234. <https://doi.org/10.1121/1.4817926>
- 1286 89. Zelazo, P.D., Anderson, J.E., Richler, J., Wallner-Allen, K., Beaumont, J.L. and Weintraub, S.
1287 (2013), II. NIH TOOLBOX COGNITION BATTERY (CB): MEASURING EXECUTIVE
1288 FUNCTION AND ATTENTION. *Monographs Society Res Child*, 78: 16-33.
1289 <https://doi.org/10.1111/mono.12032>
- 1290
- 1291

1292 **Legend**

1293

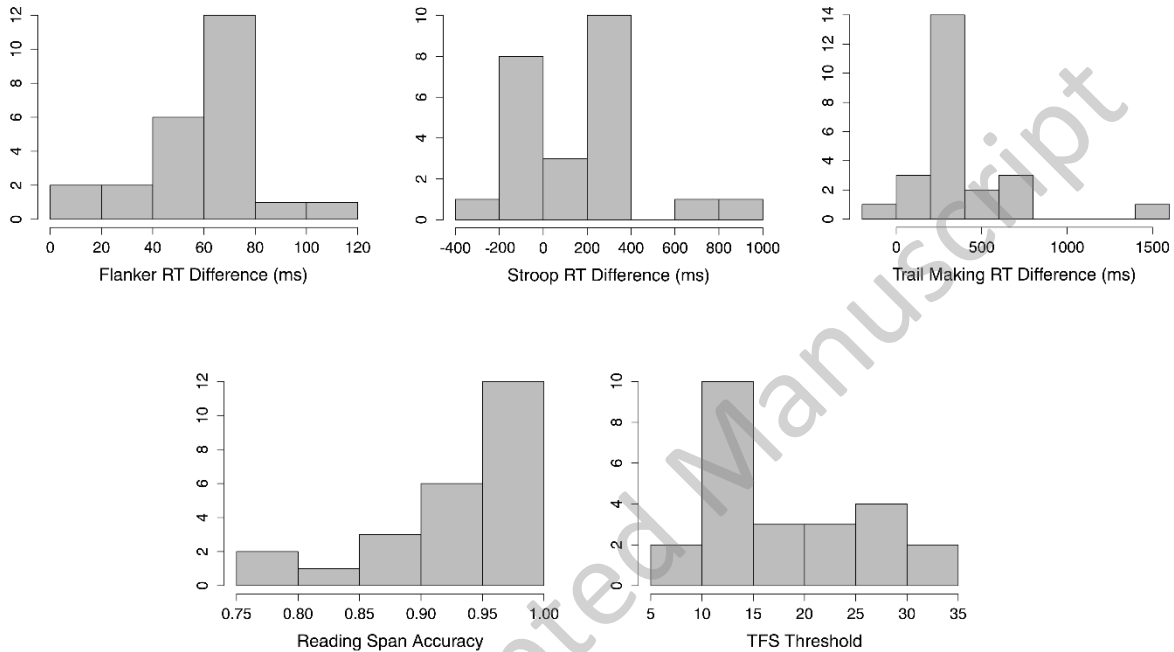
1294 **Figure 1: Continuous multi-talker spatial attention task procedure.** The figure illustrates the
 1295 visual cues presented on screen to participants during the task while they listened to a series of
 1296 short stories. The target and masker (when present) alternate between spatial locations. Only the
 1297 target story is present in the mono-talker condition, but the spatial location of the talker still
 1298 switches sides as shown by the red line. Participants are instructed to focus on the side of the
 1299 screen indicated by the “+” symbol, which marks the location of the target story. They are also
 1300 instructed to press the spacebar when they hear a color word spoken by the target speaker (red
 1301 “X” symbols) and ignore any color words in the masker story (blue “O” symbols).



1302

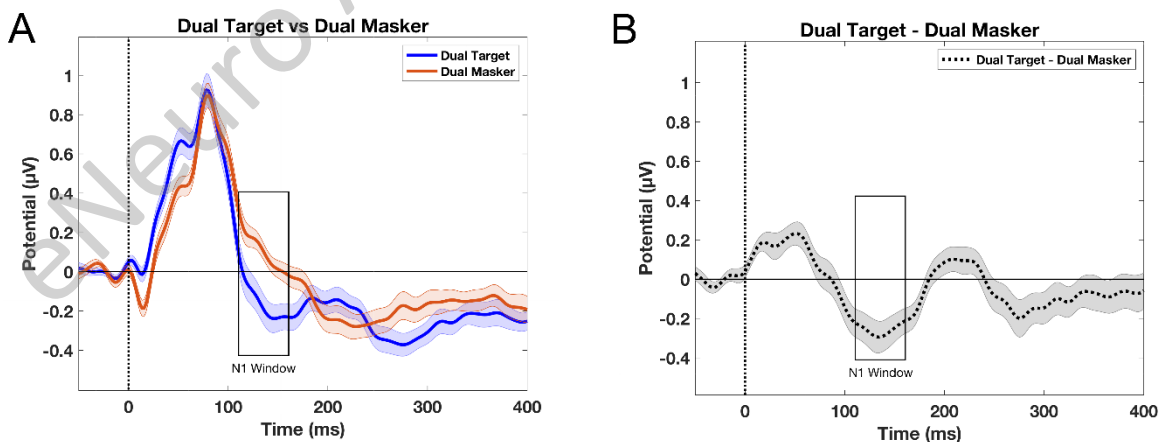
1303

1304 **Figure 2: Histograms illustrating the distribution of performance scores across five**
 1305 **cognitive and psychoacoustic predictors.** Larger RT difference scores (Flanker, Stroop, and
 1306 Trail Making) indicate greater response (cognitive) interference in each domain (i.e., slower RTs
 1307 for the incongruent/alternating trials than the congruent/sequential trials).



1308

1309 **Figure 3: Attentional modulation of speech-evoked neural activity in dual-talker conditions**
 1310 **at electrode Fz.** Error bars indicate standard error. A) Grand average ERP waveforms of the
 1311 target and masker speech. B) Difference waveforms between target and masker stories.

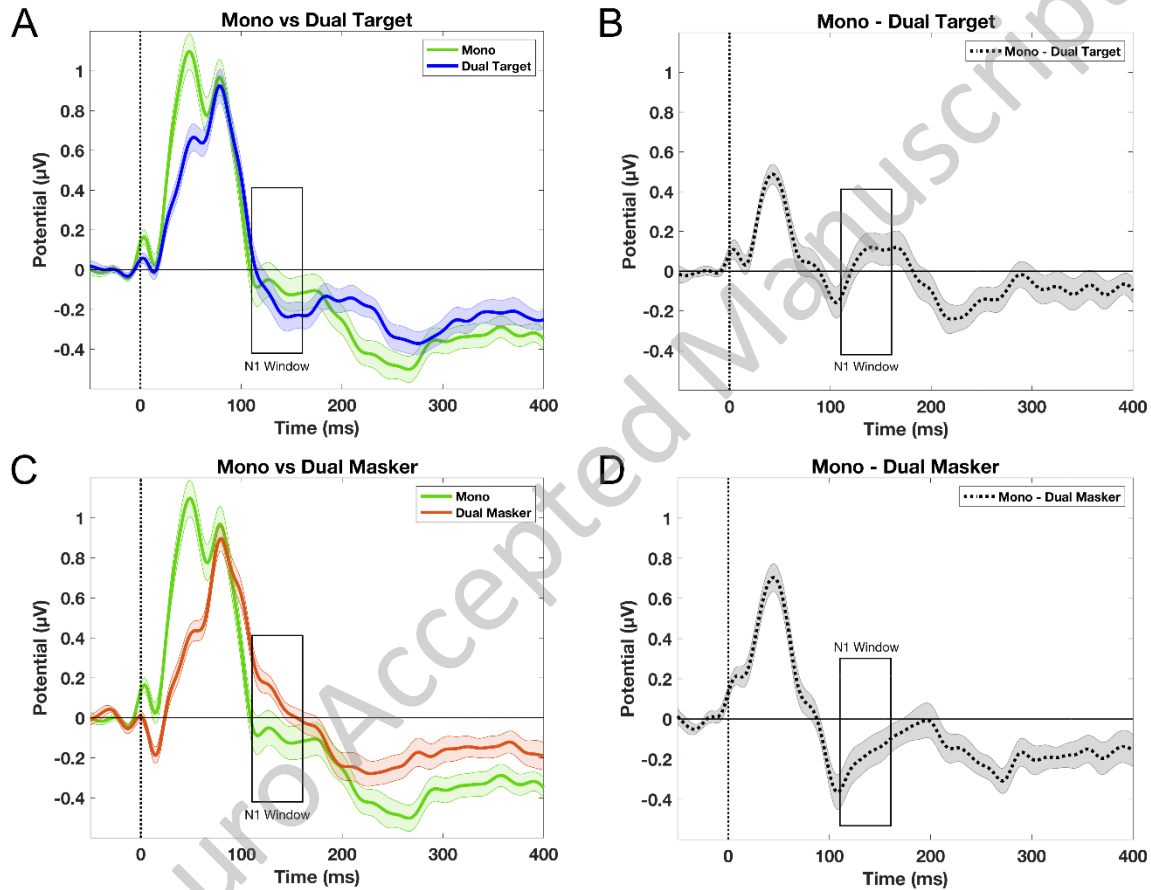


1312

1313

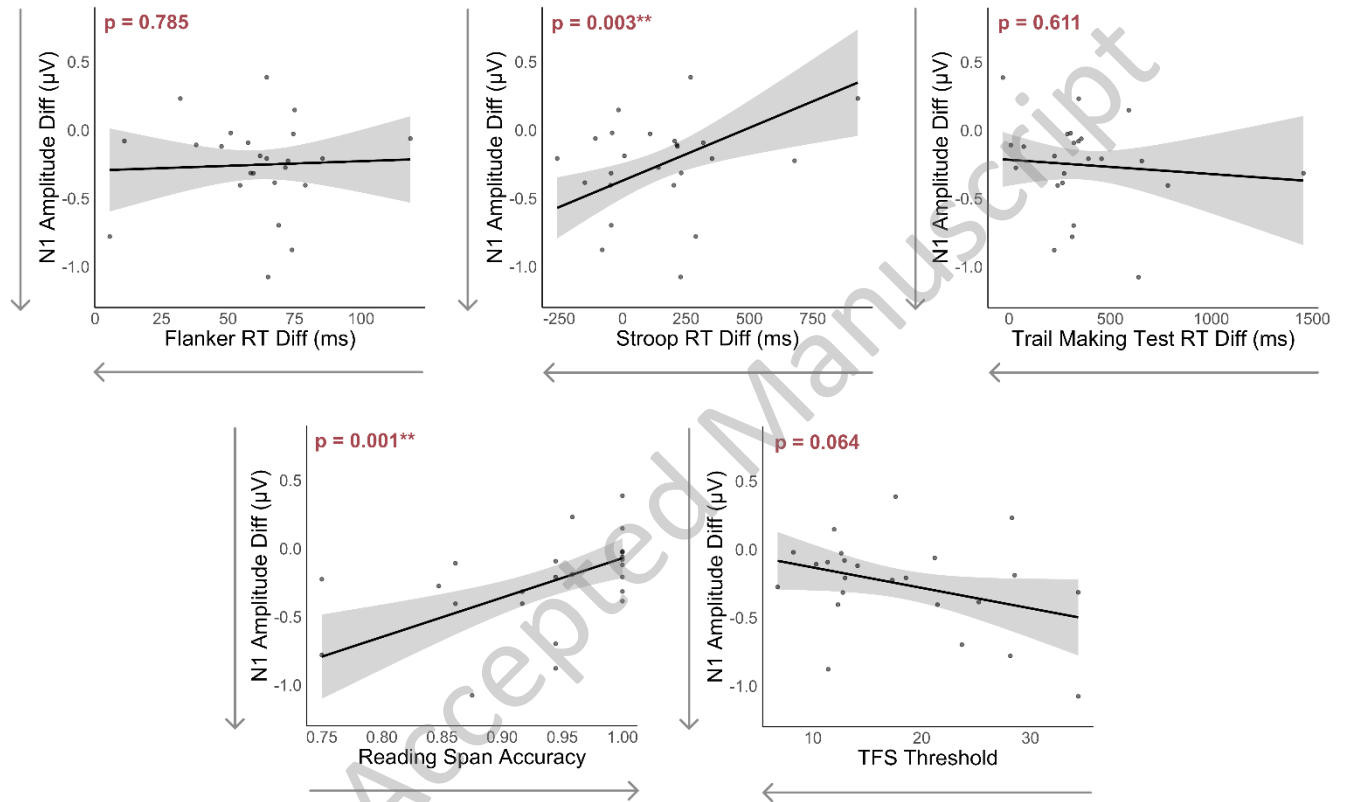
1314
1315
1316
1317
1318
1319
1320
1321

Figure 4: Attentional modulation of speech-evoked neural activity comparing mono and dual-talker conditions at electrode Fz. Error bars indicate standard error. A) Grand average ERP waveforms comparing the dual-talker target condition to the mono condition. B) Difference waveforms (mono - target), where positive values indicate larger N1 amplitudes in the dual-talker target condition than in the mono condition. C) Grand average ERP waveforms comparing the dual-talker masker condition to the mono condition. D) Difference waveforms (mono - masker speech), where negative values indicate smaller N1 amplitudes in the dual-talker masker condition compared to the mono condition.



1322
1323

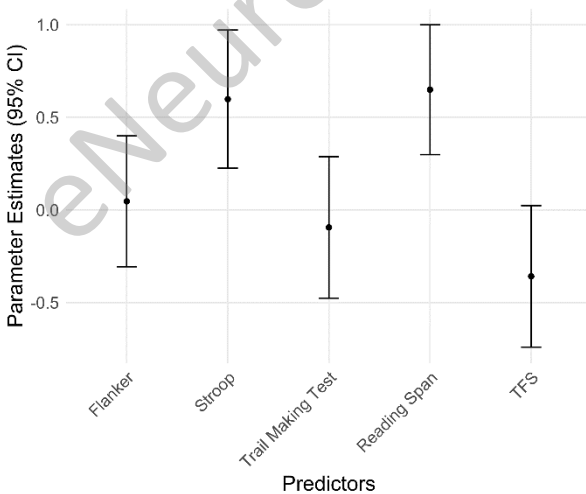
1324 **Figure 5: Effect plots illustrating the relationship between each predictor and the N1**
 1325 **amplitude difference.** All other predictors are held constant. The shaded regions represent 95%
 1326 confidence intervals for the N1 amplitude differences. N1 amplitude differences are calculated as
 1327 the N1 amplitude for the target talker minus that of the masker talker; lower values on the y-axis
 1328 correspond to larger differences in N1 amplitude between the target and masker streams. Arrows
 1329 on the x-axis denote better performance on the cognitive and psychoacoustic tests.



1330

1331

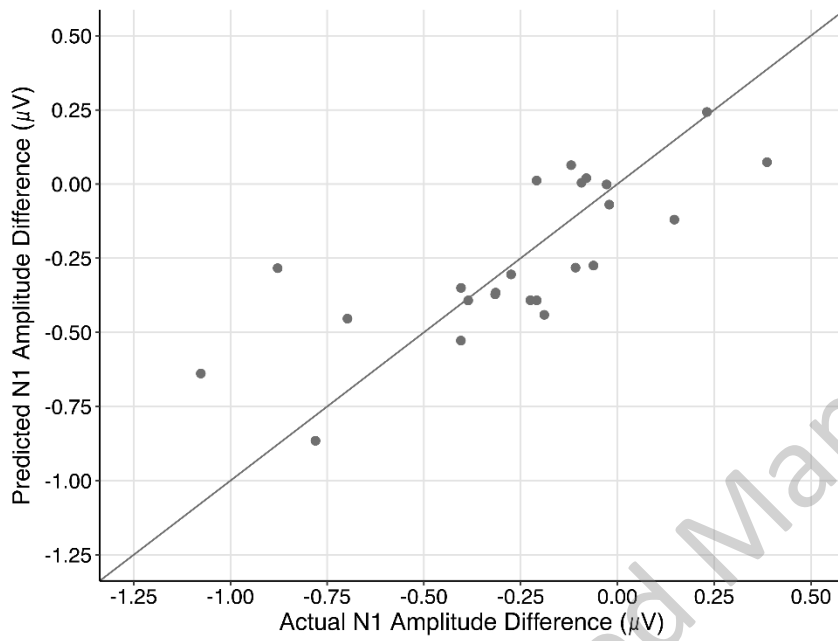
Figure 6: 95% confidence intervals for the regression coefficients of each predictor.



1332

1333
1334
1335

Figure 7: Cognitive and psychoacoustic abilities predict attentional modulation of speech representations. Results of the LOOCV. The graph shows the predicted and actual N1 amplitude differences for each participant.



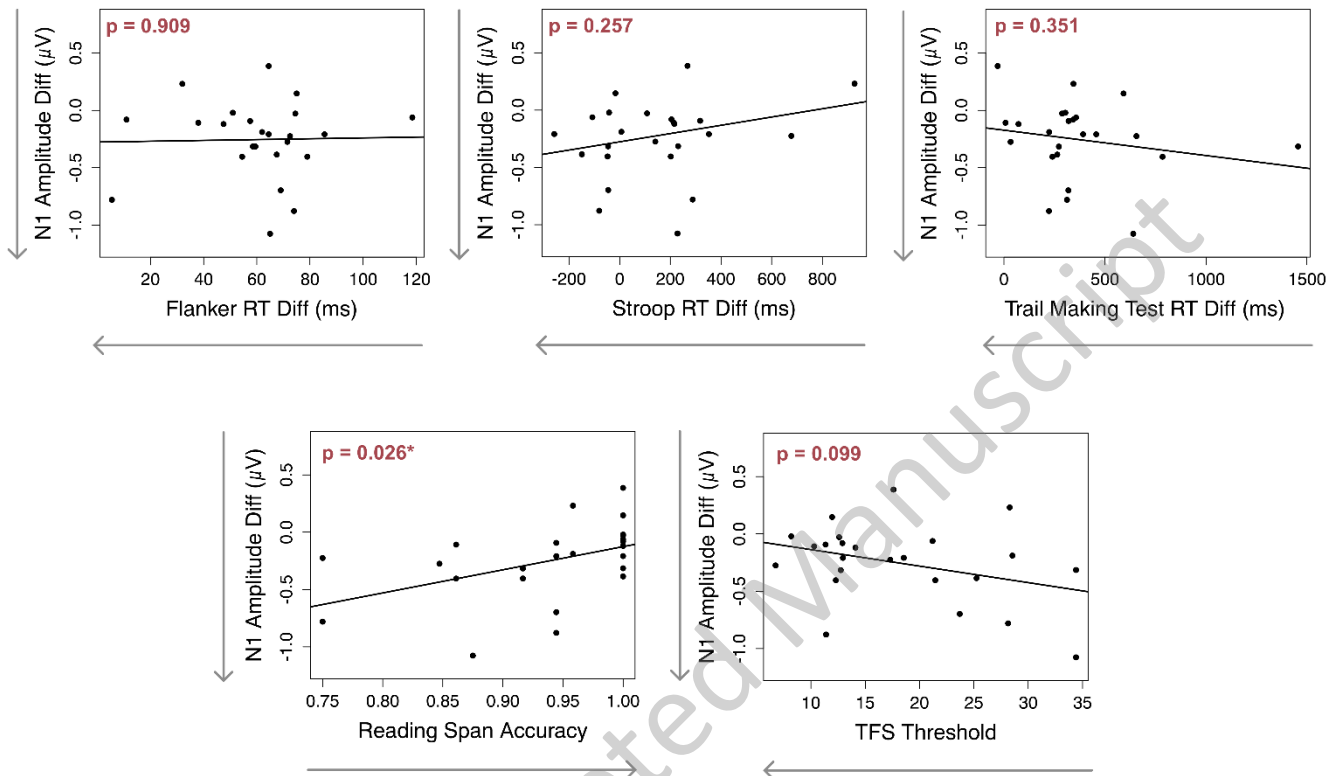
1336
1337
1338

1339

Figure 8: Bivariate correlations between each predictor and the N1 amplitude differences.

1340

Arrows on the x-axis denote better performance on the cognitive and psychoacoustic tests.



1341

1342

1343

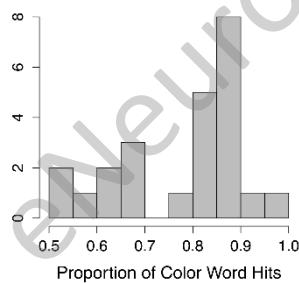
Figure 9: Histogram illustrating the distribution of proportion of color word hits in the

1344

continuous multi-talker spatial attention task. Arrows on the x-axis denote better performance

1345

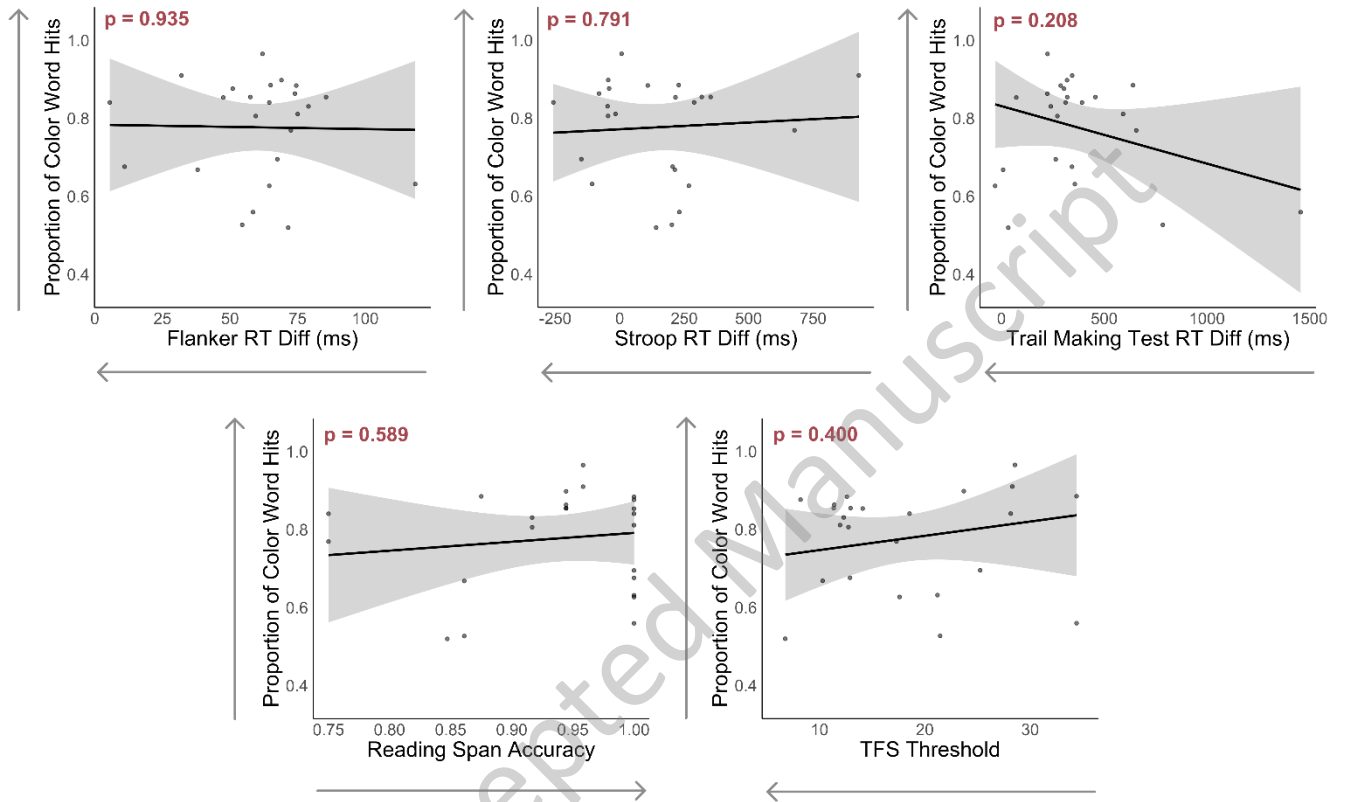
on the cognitive and psychoacoustic tests.



1346

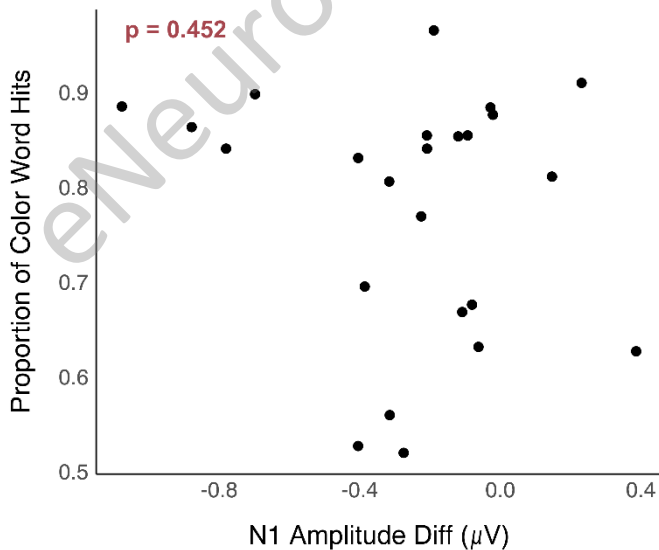
1347

1348 **Figure 10: Effect plots illustrating the relationship between each predictor and the**
 1349 **proportion of color word hits.** All other predictors are held constant. The shaded regions
 1350 represent 95% confidence intervals for the proportion of color word hits.



1351

1352 **Figure 11: Scatter plot showing the relationship between the proportion of color word hits**
 1353 **and N1 amplitude differences.** Arrows on the x-axis denote better performance on the cognitive
 1354 and psychoacoustic tests.



1355

1357 **Table 1: Results of the multiple regression analysis.** Criterion variable: N1 amplitude
 1358 difference between target and masker stories in the continuous multi-talker spatial attention task.
 1359 Predictors: performance on (1) Flanker, (2) Stroop, (3) Trail Making Test, (4) TFS, and (5)
 1360 Reading Span. All variables were z standardized. GDW (General Dominance Weight) indicates
 1361 the relative importance of each predictor in the model. Overall adjusted $R^2 = 0.467$, $p = 0.005$.

1362

	β	SE	t	p	GDW
Intercept	0.000	0.149	0.000	1.000	
Flanker	0.047	0.168	0.277	0.784	0.005
Stroop	0.598	0.178	3.370	0.003	0.166
Trail Making Test	-0.009	0.182	-0.518	0.611	0.026
Reading Span	0.649	0.167	3.894	0.001	0.278
TFS	-0.359	0.182	-1.971	0.064	0.108

1363

1364

1365

1366 **Table 2: Pairwise Pearson correlation coefficients of the performance on cognitive tasks.**

1367

	Flanker	Stroop	Trail Making Test	Reading Span	TFS
Flanker	$r = 1.000$	$r = -0.369$ $p = 0.076$	$r = 0.065$ $p = 0.763$	$r = 0.245$ $p = 0.248$	$r = -0.127$ $p = 0.555$
Stroop		$r = 1.000$	$r = 0.156$ $p = 0.467$	$r = -0.393$ $p = 0.057$	$r = 0.196$ $p = 0.358$
Trail Making Test			$r = 1.000$	$r = -0.023$ $p = 0.917$	$r = 0.520$ $p = 0.009$
Reading Span				$r = 1.000$	$r = -0.075$ $p = 0.727$
TFS					$r = 1.000$

1368

1369

1370

1371 **Table 3: Results of the second multiple regression analysis.** This model is using the same five
 1372 cognitive and psychoacoustic predictors to predict the proportion of color word hits in the
 1373 continuous multi-talker spatial attention task. All variables were z standardized. Overall adjusted
 1374 $R^2 = -0.144$, $p = 0.828$.

1375

1376

	β	SE	t	p	GDW
Intercept	0.000	0.218	0.000	1.000	0.000
Flanker	-0.020	0.246	-0.082	0.935	0.003
Stroop	0.070	0.260	0.269	0.791	0.002
Trail Making Test	-0.348	0.267	-1.305	0.208	0.067
Reading Span	0.135	0.244	0.551	0.589	0.009
TFS	0.230	0.267	0.861	0.400	0.020

1377

1378