

1 **Auditory grouping ability predicts speech-in-noise performance in cochlear implants**

2 Abbreviated title: Predictors of CI speech-in-noise

3 **Inyong Choi^{1,2*}, Phillip E. Gander³, Joel I. Berger³, Jean Hong², Sarah Colby⁴, Bob**

4 **McMurray^{1,2,4}, and Timothy D. Griffiths⁵**

5 ¹ Dept. Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242, USA

6 ² Dept. Otolaryngology – Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa

7 City, IA 52242, USA

8 ³ Dept. Neurosurgery, University of Iowa Hospitals and Clinics, Iowa City, IA 52242, USA

9 ⁴ Dept. Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA

10 ⁵ Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

11 * Corresponding author at: 250 Hawkins Dr., Iowa City, IA 52242, USA. Email address: inyong-choi@uiowa.edu

13 Number of pages: 14

14 Number of figures: 4

15 Number of tables: 1

16 Number of words

17 *Abstract:* 263

18 *Introduction:* 576

19 *Discussion:* 275

20 **Author contributions:**

21 PG, JB, TG, SC, BM, and IC designed the experiments. JH and IC ran the experiments. IC, PG,

22 JB, JH, BM, and TG analyzed and interpreted data. IC, PG, and JH prepared the manuscript,

23 with revisions and suggestions from BM and TG.

24 **Acknowledgments**

25 This work was supported by NIDCD P50 (DC000242 31) awarded to Gantz, Griffiths, and

26 McMurray, Department of Defense Hearing Restoration and Rehabilitation Program grant

27 awarded to Choi (W81XWH1910637), and DC008089 awarded to McMurray. The authors

28 declare no competing financial interests.

29

30 **Abstract**

31 *Objectives:*

32 Cochlear implant (CI) users exhibit a large variance in understanding speech in noise (SiN).

33 Past works in CI users found that spectral and temporal resolutions correlate with the SiN

34 ability, but a large portion of variance has been remaining unexplained. Our group's recent work

35 on normal-hearing listeners showed that the ability of grouping temporally coherent tones in a

36 complex auditory scene predicts SiN ability, highlighting a central mechanism of auditory scene

37 analysis that contributes to SiN. The current study examined whether the auditory grouping

38 ability contributes to SiN understanding in CI users as well.

39 *Design:*

40 47 post-lingually deafened CI users performed multiple tasks including sentence-in-noise

41 understanding, spectral ripple discrimination, temporal modulation detection, and stochastic

42 figure-ground task in which listeners detect temporally coherent tone pips in the cloud of many

43 tone pips that rise at random times at random frequencies. Accuracies from the latter three

44 tasks were used as predictor variables while the sentence-in-noise performance was used as

45 the dependent variable in a multiple linear regression analysis.

46 *Results:*

47 No co-linearity was found between any predictor variables. All the three predictors exhibited

48 significant contribution in the multiple linear regression model, indicating that the ability to detect

49 temporal coherence in a complex auditory scene explains a further amount of variance in CI

50 users' SiN performance that was not explained by spectral and temporal resolution.

51 *Conclusions:*

52 This result indicates that the across-frequency comparison builds an important auditory

53 cognitive mechanism in CI users' SiN understanding. Clinically, this result proposes a novel

54 paradigm to reveal a source of SiN difficulty in CI users and a potential rehabilitative strategy.

55

56

Introduction

57 Although cochlear implants (CIs) have been the most successful intervention for patients
58 with severe sensorineural hearing loss, persistent variability in CI speech-perception outcomes
59 remains (Gantz et al., 2016). Although much of this variability derives from peripheral factors
60 such as the degree of current spread that affects spectral resolution (Bingabr et al., 2008)
61 and/or changes to the spectral mapping in the periphery (Hamzavi et al., 2003), established CI
62 users with similar audiometric profiles and spectral resolution still differ when confronted with
63 noise (Fetterman & Domico, 2002; Noble et al., 2009). This suggests that perceptual and
64 cognitive processes may account for differences in speech perception performance. However,
65 the neural and computational mechanisms that underlie these central processes are poorly
66 understood.

67 Studies in normal auditory systems have considered speech-in-noise (SiN)
68 understanding as a “cocktail party problem” (Cherry, 1953), a problem of extracting a target
69 sound from intermixed competing sounds. A solution for the cocktail party problem is a
70 successful auditory scene analysis (Bregman, 1994), which can be understood as a chain of
71 processes including 1) sensory encoding of acoustic dynamics, 2) grouping of acoustics
72 features to form auditory objects (Darwin, 1997), and 3) across-object competition. Under this
73 ASA framework, individual differences in SiN understanding may originate from each of the ASA
74 processes. Evidence exists of individual differences in such ASA functions including encoding of
75 suprathreshold dynamics (Ruggles et al., 2011) and auditory grouping (Holmes & Griffiths,
76 2019; Holmes et al., 2021).

77 *In cochlear implant users*, most previous works investigating the sources of outcome
78 variance have focused on the status of the peripheral auditory system, that is only related to the
79 first process above (i.e. encoding acoustic cues) (Winn et al., 2016; Won et al., 2007), leaving a
80 gap in our understanding of how the central auditory functions affect speech-in-noise perception
81 in CI users.

82 Our premise is that auditory cognitive processes, independently from peripheral
83 processing, also contribute to speech-in-noise understanding in CI users. Especially, auditory
84 object formation is a key process for target identification in crowded auditory scenes. One of the
85 most effective perceptual strategies for object formation is grouping frequency components with
86 temporal coherence (Shamma et al., 2011; Teki et al., 2016; Teki et al., 2013; Teki et al., 2011).
87 Because most CI programming schemes deliver relatively intact temporal envelope patterns,
88 theoretically, temporal coherence can be perceived by electric hearing.

89 This study aimed to promote our understanding of a relationship between the central
90 auditory functions and speech-in-noise outcomes in CI users by testing whether an auditory
91 figure-ground task, an established paradigm that tests temporal coherence detection ability,
92 explains the further amount of variance in CI outcomes in addition to spectral and temporal
93 resolution. A relatively large number (47) of post-lingually deafened CI users were recruited for
94 this study and performed the stochastic figure-ground task (SFG: (Teki et al., 2011)) in which
95 listeners detect temporally coherent tone pips in the cloud of many tone pips that rise at random
96 times at random frequencies. The same subjects also performed spectral ripple discrimination
97 and a temporal modulation detection task as well as a sentence-in-noise understanding task
98 (AzBio: (Spahr et al., 2012)). The accuracy of SFG, spectral ripple discrimination, and temporal
99 modulation detection tasks were used as three predictor variables of a multiple linear regression
100 model to predict the AzBio performance, to test the hypothesis that the performance on the SFG
101 task explains a further amount of variance in CI users' AzBio performance that is not explained
102 by spectral and temporal resolution.

103

104 Materials and Methods

105 Participants

106 Forty-seven CI users, between 20 and 79 years of age (mean = 60.9 years, SD = 12.1
107 years; median = 63.3 years; % female), were recruited from the University of Iowa Cochlear

108 Implant Research Center. All the participants were post-lingually deafened and neurologically
109 normal. The average length of device use was 39.5 months (SD = 56.8 months). The average
110 duration of deafness was 22.0 years (SD = 15.0 years). Five subjects were bilateral CI users.
111 Among the remaining subjects, 66.1% had CI in the right ear. 76.3% were Hybrid CI users (i.e.,
112 electric acoustic stimulation within the same ear). Average threshold of low-frequency (i.e., 250
113 and 500 Hz) residual acoustic hearing was 59.4 dB HL (SD = 20.5 dB HL). American English
114 was the primary language for all the participants. Most participants were tested during the same
115 day as a clinical visit in which they received an annual audiological examination and device
116 tuning. Most of these CI users were bimodal or hybrid CI users who would have some residual
117 acoustic hearing. Duration of device use was obtained from clinical records. All study
118 procedures were reviewed and approved by the local Institutional Review Board.

119

120 ***Task design and procedures***

121 All CI users performed the spectral ripple discrimination, temporal modulation detection, SFG,
122 and speech in noise (AzBio).

123

124 *Speech-in-noise: AzBio*

125 Performance on a sentence-in-noise task (AzBio; (Spahr et al., 2012)) was used as a
126 dependent variable in the later multiple linear regression analysis to predict CI individuals' SiN
127 ability. Our AzBio task was performed at +5dB SNR condition at 70dB SPL in a double-walled
128 sound booth using a sound field presentation. A loudspeaker was located 1.5 m in front of the
129 subject. A single loudspeaker was used for both speech and noise (i.e., diotic listening).
130 Subjects had to repeat the sentence they heard. Audiologists counted the number of currently
131 repeated words from the outside of the sound booth. The ratio of currently repeated words to
132 the total number of words in all the presented sentences was determined and used as an
133 individual subject's accuracy for later analyses.

134

135 *Spectral Ripple and Temporal Modulation: Stimuli*

136 The stimuli for both tasks were generated in MATLAB at the time of testing based on the
137 parameters set by the Updated Maximum-Likelihood (UML) adaptive procedure (Shen et al.,
138 2015). Starting parameters were based on pilot data from 40 CI users, and subsequent trials
139 were adaptively generated based on the predictions of the UML procedure given the subject's
140 responses. Unlike traditional tasks, the UML procedure adaptively predicts what to test to best
141 estimate an individual's psychometric function. Estimates of the psychometric function were
142 updated after every trial.

143 For the spectral ripple task, stimuli were broadband noise with sinusoidal variations
144 (ripples). The ripple peaks were evenly spaced on a log-frequency scale with the density of
145 ripples held constant at 1.25 ripples per octave. The amplitude depth of the ripples (in dB) was
146 manipulated based on the UML predictions. Two standard sounds were created with a
147 randomized starting location for the spectral peak and the oddball was created with an inverted
148 phase to be maximally distinct.

149 For the temporal modulation detection task, stimuli were a complex tone comprised of
150 component frequencies at 1515, 2350, 3485, 5045, and 6990 Hz. A sine wave was overlaid on
151 the complex tone to modulate its amplitude. The envelope frequency was 20 Hz and the depth
152 of the sinusoid was determined by UML prediction. Trials either had two modulated sounds,
153 where the oddball was unmodulated, or two unmodulated sounds, where the oddball was
154 modulated.

155 Stimuli for both tasks were 500 ms in duration and ramped with a 50 ms rise/fall. To
156 compensate for intensity differences in the modulated stimuli, root mean square values were
157 equalized and the presentation level was roved randomly, so loudness could not be used as a
158 cue.

159

160 *Spectral Ripple and Temporal Modulation: Procedure*

161 Spectral and temporal resolutions were determined by spectral ripple discrimination and
162 temporal modulation detection tasks using a 3-alternate forced-choice oddball detection
163 paradigm. The task was implemented using Psychtoolbox 3 ((Brainard, 1997); (Pelli, 1997)) in
164 MATLAB (The Mathworks). On each trial, two standard stimuli and one oddball were played in
165 random order with an ISI of 750 ms. A numbered box appeared on the computer screen as
166 each stimulus played. Subjects were instructed to choose the token that differed from the other
167 two. The UML approach allowed the tasks to be much shorter than traditional staircase
168 measures; each task was 70 trials. Both tasks began with 4 practice trials to familiarize the
169 subject with the task and correct/incorrect feedback was given on every trial.

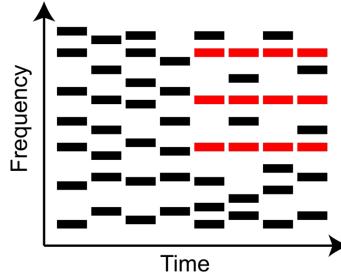
170

171 *SFG: stimuli*

172 The SFG stimuli were generated using the same design principle described in (Teki et
173 al., 2011); figure and ground tone pips overlap in time and frequency space, and they can be
174 distinguished only by their fluctuation statistics. In 50% of the trials, the tonal components
175 repeated in frequency over 2 seconds, which popped out as “figure” out of “ground.”

176 In our experiment, the frequencies of the figure varied from trial to trial but the number of
177 figure tone pips was always six. To avoid subjects’ spectral resolution or low-frequency residual
178 acoustic hearing confounding their SFG ability, the frequency gap between adjacent figure tone-
179 pips (i.e., repeated tonal components) was at least one octave and they were all above 1kHz.
180 Figure 1 shows an example schematic stimulus containing the “figure” tone pips in the second
181 half.

182



183

184 **Figure 1.** A schematic spectrogram of SFG stimulus.

185 All stimuli were created using MATLAB software (The Mathworks) at a sampling rate of
186 44.1 kHz and 16-bit resolution.

187

188 *SFG: task*

189 The SFG task was implemented in custom-written Matlab scripts (The Mathworks) using
190 the Psychtoolbox 3 toolbox ((Brainard, 1997); (Pelli, 1997)). The SFG was conducted in an
191 acoustically-treated, electrically-shielded booth with a single loudspeaker (model LOFT40, JBL)
192 positioned at a 0° azimuth angle at a distance of 1.2 m. Visual stimuli were presented via a
193 computer monitor located 0.5m in front of the subject at eye level. Sound levels were the same
194 across subjects.

195 For each trial, participants saw a fixation cross on the computer screen. r a response
196 was given.

197

198 **Statistical analyses**

199 We related each predictor to speech perception performance on AzBio. The final model
200 is given in (1), in the syntax of the linear modeling (i.e., lm()) function in R.

201

202 *Speech Perception ~ 1 + SFG + SpecRipple + TempModul* (1)

203

204

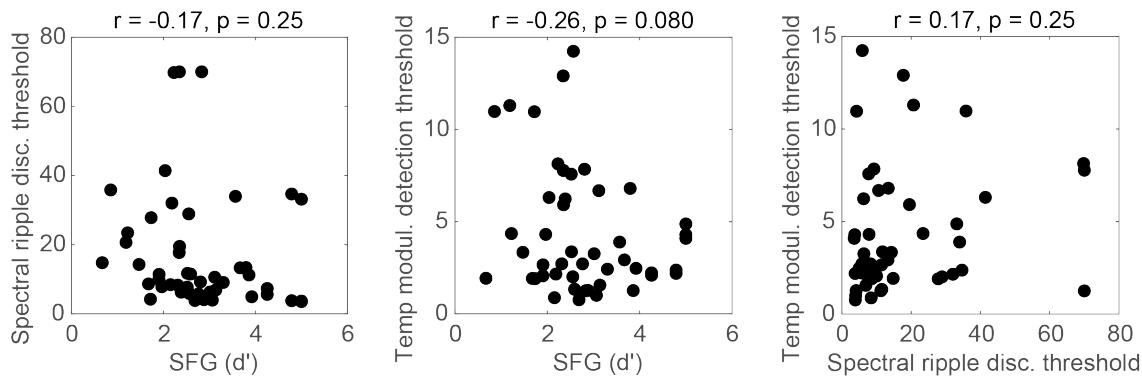
205

Results

206 **Evaluation of independent variables in bivariate analyses**

207 We started by evaluating the correlations among all the independent variables: to check
208 for co-linearity prior to multiple linear regression analysis. No significant correlation was found
209 between any predictor variables. The relationship between the predictor variables is shown in
210 Figure 2 as scatter plots.

211



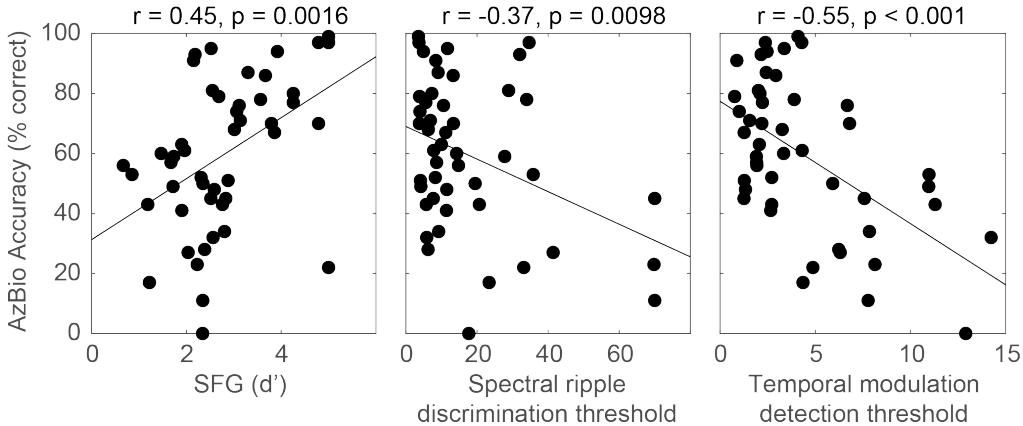
212

213 **Figure 2.** Results from predictor co-linearity analysis. No significant correlation is observed.

214

215 Next, we conducted exploratory bivariate analyses examining correlations between each
216 independent variable and AzBio accuracy. All three predictors exhibited a statistically significant
217 correlation with SiN ability. These are shown in Figure 3. In Figure 3, the spectral ripple
218 discrimination and temporal modulation detection threshold values are in an arbitrary unit,
219 measured as crossover during an adaptive procedure.

220



221

222 **Figure 3.** Results from bivariate correlation analyses. Spectral ripple
223 modulation detection threshold values are in an arbitrary unit.

224

225 ***Multiple linear regression***

226 Following bivariate analyses, we conducted a multiple linear regression analysis to
227 determine which of the independent variables predicted AzBio accuracy when accounting for all
228 others (see Table 1 in its entirety). When adjusted for the number of independent variables, the
229 model accounted for 46.3% of the variance in AzBio accuracy, $F(3, 43) = 12.4, p < 0.00001$,
230 Adjusted $R^2 = 0.426$. All three predictors reached statistical significance.

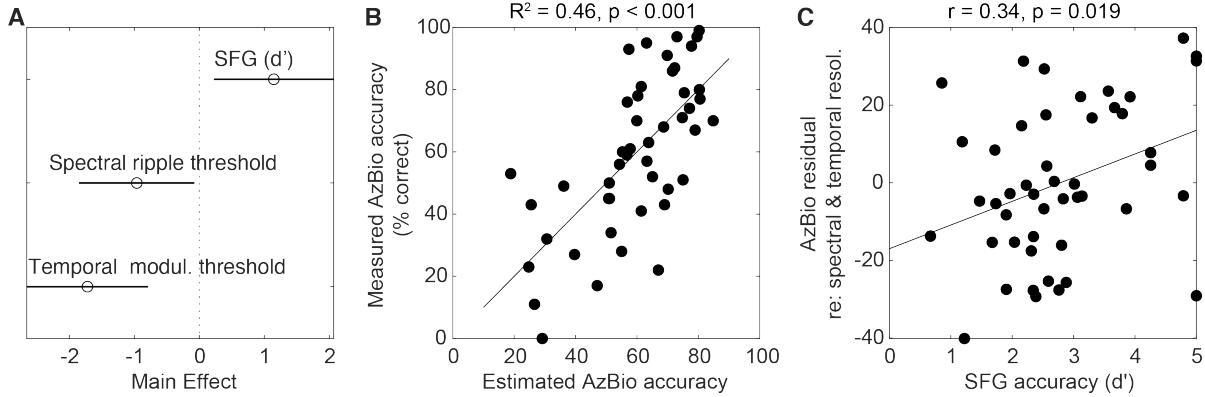
231

232 **Table 1:** Results from multiple linear regression on SiN accuracy (N=47, R²=0.463):

AzBio	β (normalized)	SE	T(43)	p	Partial ρ
SFG	0.292	2.65	2.51	0.0160	0.357
SpecRippleThres	-0.250	0.167	-2.18	0.0345	-0.316
TempModulThres	-0.434	0.863	-3.72	<0.001	-0.493

233

234



235 **Figure 4.** Results from multiple linear regression analysis. **A.** Main effects of predictor variables. **B.**
236 Relationship between estimated AzBio accuracy (i.e., the model output) and measured AzBio accuracy
237 (i.e., the dependent variable). **C.** Relationship between SFG accuracy and the residual of AzBio accuracy
238 after regressing out the other two predictor variables (i.e., spectral and temporal resolution).

239

240

Discussion

241 In this study, post-lingually deafened CI users performed a stochastic figure-ground (SFG)
242 task in which listeners detect temporally coherent frequency components in the cloud of tone
243 pips that rise at random times at random frequencies. The bivariate correlation between figure-
244 detection performance (d' -prime) and sentence-in noise performance (AzBio score) was high as
245 $r = 0.45$ ($p < 0.005$). The effect size is greater than that in normal hearing subjects reported in
246 (Holmes & Griffiths, 2019). Moreover, multiple linear regression demonstrated a significant
247 effect of figure detection (normalized beta coefficient = 0.29, $p < 0.05$) even after accounting for
248 the fidelity of spectral and temporal encoding in the auditory periphery. The combined model
249 explained 46% of the variance in SiN performance. This work has therefore established a
250 relationship between a simple measure of the cross-frequency grouping of electrically coded
251 signals, relevant to SiN ability.

252 This result suggests that auditory-cognitive mechanisms play a powerful explanatory in CI
253 users' outcomes in everyday communications. Adopting the SFG task in clinics may reveal a

254 source of SiN difficulty in CI users that has not been recognized in the current clinical practice,
255 which potentially opens up an opportunity for a novel aural rehabilitative strategy for CI users.

256 The figure detection ability during the SFG task is unlikely the only auditory cognitive
257 mechanism that contributes to SiN performance. Although 47 was a relatively large sample size
258 for a CI study, the number of predictor variables was limited to three to ensure a reasonable
259 statistical power. A future study will consider more auditory-cognitive mechanisms (e.g., auditory
260 working memory: (Akeroyd, 2008; Dryden et al., 2017; Kim et al., 2020)) as well as linguistic
261 and general cognitive mechanisms.

262

263

264

References

265

- 266 Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual
267 differences in cognitive ability? A survey of twenty experimental studies with normal and
268 hearing-impaired adults. *Int J Audiol*, 47 Suppl 2, S53-71.
269 <https://doi.org/10.1080/14992020802301142>
- 270 Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of
271 excitation in cochlear implants. *Hear Res*, 241(1-2), 73-79.
272 <https://doi.org/10.1016/j.heares.2008.04.012>
- 273 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10(4), 433-436.
274 <https://www.ncbi.nlm.nih.gov/pubmed/9176952>
- 275 Bregman, A. S. (1994). *Auditory scene analysis : the perceptual organization of sound*
276 (Paperback edition. ed.). MIT Press.
- 277 Cherry, C. E. (1953). *Some experiments on the recognition of speech, with one and with two*
278 *ears*. American Institute of Physics for the Acoustical Society of America.
- 279 Darwin, C. J. (1997). Auditory grouping. *Trends Cogn Sci*, 1(9), 327-333.
280 [https://doi.org/10.1016/S1364-6613\(97\)01097-8](https://doi.org/10.1016/S1364-6613(97)01097-8)
- 281 Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between
282 Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A
283 Systematic Literature Review and Meta-Analysis. *Trends Hear*, 21, 2331216517744675.
284 <https://doi.org/10.1177/2331216517744675>
- 285 Fetterman, B. L., & Domico, E. H. (2002). Speech Recognition in Background Noise of Cochlear
286 Implant Patients. *Otolaryngology -- Head and Neck Surgery*, 126(3), 257-263.
287 <https://doi.org/10.1067/mhn.2002.123044>
- 288 Gantz, B. J., Dunn, C., Oleson, J., Hansen, M., Parkinson, A., & Turner, C. (2016). Multicenter
289 clinical trial of the Nucleus Hybrid S8 cochlear implant: Final outcomes. *The*
290 *Laryngoscope*, 126(4), 962-973.
- 291 Hamzavi, J., Baumgartner, W.-d., Pok, S. M., Franz, P., & Gstoeftner, W. (2003). Variables
292 affecting speech perception in postlingually deaf adults following cochlear implantation.
293 *Acta oto-laryngologica*, 123(4), 493-498.
- 294 Holmes, E., & Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory
295 grouping processes predict difficulties with speech-in-noise perception. *Sci Rep*, 9(1),
296 16771. <https://doi.org/10.1038/s41598-019-53353-5>
- 297 Holmes, E., Zeidman, P., Friston, K. J., & Griffiths, T. D. (2021). Difficulties with Speech-in-
298 Noise Perception Related to Fundamental Grouping Processes in Auditory Cortex.
299 *Cereb Cortex*, 31(3), 1582-1596. <https://doi.org/10.1093/cercor/bhaa311>
- 300 Kim, S., Choi, I., Schwalje, A. T., Kim, K., & Lee, J. H. (2020). Auditory Working Memory
301 Explains Variance in Speech Recognition in Older Listeners Under Adverse Listening
302 Conditions. *Clin Interv Aging*, 15, 395-406. <https://doi.org/10.2147/cia.S241976>
- 303 Noble, W., Tyler, R. S., Dunn, C. C., & Bhullar, N. (2009). Younger-and older-age adults with
304 unilateral and bilateral cochlear implants: speech and spatial hearing self-ratings and
305 performance. *Otology & neurotology: official publication of the American Otological*
306 *Society, American Neurotology Society [and] European Academy of Otology and*
307 *Neurotology*, 30(7), 921.
- 308 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers
309 into movies. *Spat Vis*, 10(4), 437-442. <https://www.ncbi.nlm.nih.gov/pubmed/9176953>
- 310 Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2011). Normal hearing is not enough
311 to guarantee robust encoding of suprathreshold features important in everyday

- 312 communication. *Proc Natl Acad Sci U S A*, 108(37), 15516-15521.
313 <https://doi.org/10.1073/pnas.1108912108>
- 314 Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory
315 scene analysis. *Trends Neurosci*, 34(3), 114-123.
316 <https://doi.org/10.1016/j.tins.2010.11.002>
- 317 Shen, Y., Dai, W., & Richards, V. M. (2015). A MATLAB toolbox for the efficient estimation of
318 the psychometric function using the updated maximum-likelihood adaptive procedure.
319 *Behav Res Methods*, 47(1), 13-26. <https://doi.org/10.3758/s13428-014-0450-6>
- 320 Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L.
321 M., Oakes, T., & Cook, S. (2012). Development and validation of the AzBio sentence
322 lists. *Ear Hear*, 33(1), 112-117. <https://doi.org/10.1097/AUD.0b013e31822c2549>
- 323 Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T. D., & Chait, M. (2016). Neural
324 Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence.
325 *Cereb Cortex*, 26(9), 3669-3680. <https://doi.org/10.1093/cercor/bhw173>
- 326 Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex
327 acoustic scenes based on temporal coherence. *eLife*, 2, e00699.
328 <https://doi.org/10.7554/eLife.00699>
- 329 Teki, S., Chait, M., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2011). Brain bases for
330 auditory stimulus-driven figure-ground segregation. *J Neurosci*, 31(1), 164-171.
331 <https://doi.org/10.1523/jneurosci.3788-10.2011>
- 332 Winn, M. B., Won, J. H., & Moon, I. J. (2016). Assessment of Spectral and Temporal Resolution
333 in Cochlear Implant Users Using Psychoacoustic Discrimination and Speech Cue
334 Categorization. *Ear Hear*, 37(6), e377-e390.
335 <https://doi.org/10.1097/AUD.0000000000000328>
- 336 Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with
337 speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol*, 8(3), 384-
338 392. <https://doi.org/10.1007/s10162-007-0085-8>
- 339