



Knight, Sarah and Heinrich, Antje (2017) Different measures of auditory and visual stroop interference and their relationship to speech intelligibility in noise. *Frontiers in Psychology*, 8 . ISSN 1664-1078

Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/41296/1/KnightHeinrich_2017.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see: <http://creativecommons.org/licenses/by/2.5/>

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Different Measures of Auditory and Visual Stroop Interference and Their Relationship to Speech Intelligibility in Noise

Sarah Knight^{1*}, Antje Heinrich¹

¹MRC Institute of Hearing Research, University of Nottingham, United Kingdom

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Auditory Cognitive Neuroscience

ISSN:
1664-1078

Article type:
Original Research Article

Received on:
18 Oct 2016

Accepted on:
06 Feb 2017

Provisional PDF published on:
06 Feb 2017

Frontiers website link:
www.frontiersin.org

Citation:
Knight S and Heinrich A(2017) Different Measures of Auditory and Visual Stroop Interference and Their Relationship to Speech Intelligibility in Noise. *Front. Psychol.* 8:230.
doi:10.3389/fpsyg.2017.00230

Copyright statement:
© 2017 Knight and Heinrich. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

1 **Different Measures of Auditory and Visual Stroop Interference**
2 **and Their Relationship to Speech Intelligibility in Noise**

3
4 Sarah Knight^{1*}, Antje Heinrich¹

5
6 ¹Medical Research Council Institute of Hearing Research, University of Nottingham,
7 Nottingham, UK.

8
9 ***Correspondence:**

10 Sarah Knight

11 slk34@cantab.net

12
13 **Keywords: speech-in-noise, inhibition, aging, Stroop tasks, scoring**

Provisional

14 **Abstract**

15 Inhibition – the ability to suppress goal-irrelevant information – is thought to be an important
16 cognitive skill in many situations, including speech-in-noise (SiN) perception. One way to
17 measure inhibition is by means of Stroop tasks, in which one stimulus dimension must be
18 named while a second, more prepotent dimension is ignored. The to-be-ignored dimension
19 may be relevant or irrelevant to the target dimension, and the inhibition measure – Stroop
20 interference (SI) – is calculated as the reaction time difference between the relevant and
21 irrelevant conditions. Both SiN perception and inhibition are suggested to worsen with age,
22 yet attempts to connect age-related declines in these two abilities have produced mixed
23 results. We suggest that the inconsistencies between studies may be due to methodological
24 issues surrounding the use of Stroop tasks. First, the relationship between SI and SiN
25 perception may differ depending on the modality of the Stroop task; second, the traditional SI
26 measure may not account for generalized slowing or sensory declines, and thus may not
27 provide a pure interference measure.

28 We investigated both claims in a group of 50 older adults, who performed two Stroop tasks
29 (visual and auditory) and two SiN perception tasks. For each Stroop task, we calculated
30 interference scores using both the traditional difference measure and methods designed to
31 address its various problems, and compared the ability of these different scoring methods to
32 predict SiN performance, alone and in combination with hearing ability. Results from the two
33 Stroop tasks were uncorrelated and had different relationships to SiN perception. Changing
34 the scoring method altered the nature of the predictive relationship between Stroop scores and
35 SiN perception, which was additionally influenced by hearing ability. These findings raise
36 questions about the extent to which different Stroop tasks and/or scoring methods measure
37 the same aspect of cognition. They also highlight the importance of considering additional
38 variables such as hearing ability when analysing cognitive variables.

39 1 Introduction

40 Inhibition – the ability to suppress goal-irrelevant information (MacLeod, 1991) – is thought
41 to be important in many situations. One of these situations is speech-in-noise (SiN)
42 perception, in which listeners aim to focus on the foreground (target speech) and ignore the
43 background (distractor) sound. The ability to inhibit irrelevant information has been
44 suggested to worsen with age (Hasher & Zacks, 1988), with implications across a variety of
45 cognitive domains including language, memory and attention (Burke, 1997; Stoltzfus, Hasher
46 & Zacks, 1996). This cognitive decline has potential consequences for everyday activities
47 such as reading and text comprehension (Dywan & Murphy, 1996) and even engaging in
48 appropriate social behaviour (von Hippel, 2007). The ability to understand speech-in-noise is
49 also observed to worsen with age, affecting the ability to hold conversations and engage in
50 social activities (CHABA, 1988). Given the suggested importance of inhibition for SiN
51 perception, researchers have begun to ask whether or not age-related declines in inhibition
52 could account, at least in part, for the observed difficulties older adults have when listening in
53 noisy environments. However, answering this question has been made difficult by the fact
54 that it is not clear what role modality plays in the measurement of inhibition (whether or not
55 inhibition tasks in different modalities measure the same underlying ability) and whether the
56 standard scoring method adequately accounts for other, unconnected, age-related changes.

57 In the following section we introduce two types of Stroop task, a paradigm commonly used to
58 assess inhibitory abilities and the focus of this study. We first explain the nature of Stroop
59 tasks, and discuss the effect perceptual modality has on task outcomes. Next, we explore the
60 effect of age-related changes on Stroop interference and consider potential underlying
61 mechanisms. Finally, we discuss the relationship between the most common outcome,
62 measure of Stroop interference, reaction times (RTs), and strength of inhibition, and propose
63 that trials which are responded to more slowly may not only represent inhibition more
64 accurately than trials responded to more quickly but may also better reveal differential levels
65 of inhibition between participants. We then turn to speech-in-noise perception, and discuss
66 the possible role of inhibition in SiN perception. In particular, we focus on the role inhibition
67 plays during lexical access, a key element of speech perception, and consider how changes
68 across the lifespan in lexical access effects might indicate age-related changes in inhibition.
69 Finally, we discuss the results obtained from existing studies designed to test the relationship
70 between inhibition and SiN perception, and suggest some reasons why these discrepancies
71 might arise.

72 1.1 Stroop tasks

73 One common means of assessing inhibition is by using variants of the Stroop task (Stroop,
74 1935). In the traditional visual colour-word Stroop task (ibid.), participants are required to
75 name the ink colour of a string of letters, irrespective of the letters themselves. The string of
76 letters can be either meaningless (e.g. XXXX) – the neutral condition – or can form a
77 conflicting colour word (e.g. BLUE printed in red) – the incongruent condition. Since word
78 reading is a more prepotent response than colour naming in this situation (Melara & Algom,
79 2003), word naming has the potential to interfere with colour naming. In order to prevent this
80 interference, participants must attempt to inhibit, or suppress, the incongruent word. The
81 difference in reaction time (RT) between colour naming in the neutral condition and colour
82 naming in the incongruent condition is taken as a measure of inhibitory ability, and termed
83 Stroop interference (SI). Besides the traditional visual paradigm, auditory versions of the
84 Stroop task have also been successfully used (e.g. Green & Barber, 1981; Morgan & Brandt,
85 1989). In auditory Stroop tasks, participants are required to respond as quickly as possible to
86 some perceptual feature of a word (e.g. speaker gender, voice pitch, stimulus location) while

87 ignoring the semantic information, which can be either irrelevant (e.g. “cat”) or conflicting
88 (e.g. “man” spoken by a woman, “low” in a high-pitched voice, “right” heard in the left ear).
89 Again, SI is typically obtained by calculating the difference in reaction time between feature
90 naming with irrelevant semantic content and feature naming with an incongruent semantic
91 distractor.

92 **1.1.1 Stroop tasks across modalities.**

93 The visual and auditory versions of the Stroop task are generally assumed to tap the same
94 underlying domain-general inhibitory ability; however, the relationship between the two
95 measures and the extent to which this assumption is true remains unclear. On the one hand,
96 there is evidence to suggest that carefully-matched Stroop tasks presented across different
97 modalities do probe shared inhibitory processes, producing similar patterns of neural
98 activation and correlated behavioural responses (Roberts & Hall, 2008). On the other hand, it
99 has been shown that, even within the same modality, measures of inhibition that are not so
100 closely matched do not correlate within individuals, suggesting either that there is no single
101 inhibitory function supporting performance across different tasks and/or that task-specific
102 demands determine individual differences more strongly than general inhibitory abilities
103 (Shilling, Chetwynd & Rabbitt, 2002). This suggests that any two inhibition tasks, either
104 within or across modalities, are unlikely to be comparable unless they have been deliberately
105 matched, and in particular that an auditory Stroop task cannot automatically be assumed to be
106 an alternative way of measuring the same ability tapped by a given visual Stroop task. In the
107 current study we will address the question of the relationship between visual and auditory
108 versions of the Stroop task by comparing scores from the same participants on an auditory
109 and a visual Stroop task, both deliberately chosen to meet certain criteria.

110 **1.1.2 Age-related declines in Stroop performance.**

111 When calculated in the traditional way, SI (Stroop interference) on both visual and auditory
112 tasks is generally observed to increase with age, implying a worse performance on the
113 incongruent Stroop task compared to the neutral condition and – hence – poorer inhibition.
114 However, it has long been recognised that no task is ever a “pure” measure of a given
115 cognitive function, but instead includes other, additional processes – something referred to as
116 the “impurity principle” (Surprenant & Neath, 2009). In the case of the Stroop task, it has
117 been suggested that these age-related increases in SI could be due, at least in part, to just such
118 additional processes; that is, that there are potential confounds with non-inhibitory factors
119 created by the methods typically used to calculate SI (Ben-David & Schneider, 2009) – and
120 that methods should be used which account for these factors.

121 One of these confounds is generalised age-related slowing. In the traditional SI measure,
122 inhibition is represented by the absolute difference in time taken to name the background
123 colour between conditions with and without a distracting colour word. A change in the speed
124 of processing would slow performance on all tasks by the same factor (Verhaeghen &
125 Cerella, 2002; Cerella & Hale, 1994), leading to a proportional increase of RTs in
126 incongruent and neutral conditions; this would result in a larger absolute difference between
127 RTs in the two condition, and thus a larger SI (Shilling et al, 2002; Ben-David and Schneider,
128 2009). Crucially, in such a case the increased SI does not necessarily represent any decline in
129 inhibitory ability, but a change in processing speed. One way to address this issue is to use a
130 method for calculating Stroop scores which accounts for, or factors out, changes in overall
131 processing speed. For example, it is possible to use normalised scores, in which the RT in the
132 incongruent condition is divided by the RT in the neutral condition, thus removing any

133 changes in SI caused by proportional RT increases in both conditions. This is further
134 discussed in Sections 4.2 below.

135 While a generalised slowing of processing speed is expected to affect Stroop tasks across
136 different modalities in similar ways, the confounding effects of sensory change will be
137 specific to the perceptual domain of any given Stroop task. For visually presented Stroop
138 tasks, such confounding effects may be particularly critical when they adversely affect the RT
139 of the incongruent condition. If we accept the proposal of Melara & Algom (2003) that the
140 Stroop interference effect arises due to a failure to inhibit the more rapidly accessed printed
141 word until access to the incongruent colour name is achieved, then changes in colour vision
142 may make access to the colour word slower and/or more difficult, thereby increasing reaction
143 times during colour naming (Ben-David & Schneider, 2010). Such changes could be brought
144 about by age-related yellowing of the lens and a loss of photo receptors (Anstey et al, 2002;
145 Werner & Steele, 1988). These age-related changes in colour vision do not affect word
146 reading (Salthouse & Meinz, 1995), the speed of which remains largely unchanged with age
147 provided the words are sufficiently legible (Akutsu et al, 1991). As a result, the difference
148 between the time taken to read incongruent words and to name ink colours will be much
149 greater for individuals with an age-related decline in colour vision than for those with better
150 colour vision (i.e. younger adults). Melara & Algom (2003) characterised this discrepancy
151 between colour naming speed and reading speed as the “Dimensional Imbalance”, or DI.
152 Having a larger DI – that is, a greater discrepancy in processing time between reading and
153 colour naming – puts individuals at an increased risk of a failure of inhibition (as expressed in
154 larger SIs), since participants have to suppress the irrelevant word for longer. In this case,
155 then, increased SI scores may reflect a combination of reduced inhibitory control and an
156 increased likelihood of inhibitory failure caused by differences in processing speed for words
157 as opposed to colours (i.e. a large DI). One way to address this issue is to use a method for
158 calculating Stroop scores which accounts for, or factors out, differences in DI. For example, it
159 is possible to regress RTs in the incongruent condition on DI scores, and then use the
160 residuals as a measure of Stroop interference. This is discussed further in Section 4.2 below.

161 In the current study we will examine the effect of general age-related slowing and age-related
162 sensory changes by comparing alternative scoring methods that capture age-related changes
163 in inhibitory ability to different extents.

164 **1.1.3 RT distributions in Stroop tasks.**

165 In addition to questions of how to appropriately capture the differential age trajectories of the
166 processes contributing to the overall effect, there is a further issue with the way in which
167 Stroop scores are traditionally calculated, namely that they usually use an average score over
168 all trials. If it is true (e.g. Ridderinkhof et al, 2004) that the strength of inhibition depends on
169 the overall processing time, with the slowest responses allowing more time for inhibition to
170 build up, then differences in inhibitory ability are likely to be most evident during those trials
171 with the longest reaction times. That is, trials with longer reaction times will be more
172 informative when assessing inhibitory differences than trials with shorter reaction times,
173 since the gap between those with good inhibition and those with poor inhibition will be at its
174 most pronounced. In averaging over all trials, the traditional SI measure may blur crucial
175 information by mixing outcomes from some informative (slow) trials with outcomes from
176 many uninformative (fast) trials. In the second part of the paper we will examine this
177 hypothesis by investigating the differing extent of Stroop interference for slow and fast trials.

178 **1.2 Speech-in noise perception and inhibition**

179 Research into SiN perception difficulties in older adults has revealed that only some of these
180 difficulties can be accounted for by hearing loss, and that other abilities must play a role
181 (Schneider & Pichora-Fuller, 2000; Wingfield & Tun, 2007). One of those abilities is
182 cognition, which must be examined alongside hearing loss in order to better explain age-
183 related difficulties (Akeroyd, 2008). Cognition is not a unitary construct, and has many
184 different components. The exact number and nature of the cognitive components varies
185 across different cognitive models; however, inhibition is generally identified as a core ability
186 (e.g. Diamond, 2013; Baddeley, 2011; Conway & Eagle, 1994; Friedman & Miyake, 2004).
187 Two potential ways in which inhibition may affect SiN perception have been suggested. First,
188 poor inhibition may increase susceptibility to background noise during SiN listening (Janse,
189 2012). This implies not only that those with poor inhibition will perform worse on SiN tasks
190 than those with good inhibition, but also that their difficulties may increase
191 disproportionately as the signal-to-noise ratio (SNR) becomes more adverse. Second, it is
192 suggested that poor inhibition may make it harder for listeners to successfully select the target
193 during lexical access.

194 **1.2.1 Lexical access and inhibition.**

195 One way to conceptualise lexical access is in terms of the Neighborhood Activation Model
196 (NAM) (Luce & Pisoni, 1998). The NAM proposes that items in the mental lexicon are
197 organised into similarity neighborhoods, defined as all words that can be created from a
198 target item by adding, deleting or substituting a single phoneme. Any given target word will
199 activate both the target and, to varying degrees, its surrounding neighborhood, which may be
200 large (dense) or small (sparse); furthermore, words which are more commonly encountered
201 (have a high frequency of occurrence) will be activated more strongly than those less
202 commonly encountered. Words are therefore classified as “lexically easy” if they have a high
203 word frequency and relatively sparse neighborhoods, and as “lexically hard” if they have a
204 low word frequency and relatively dense neighborhoods. It is assumed that inhibition plays a
205 larger role in the perception of lexically hard words than easy words. It is therefore expected
206 not only that listeners will be less likely to correctly identify lexically hard words than
207 lexically easy words, but also that individual differences in inhibition will relate more closely
208 to the perception of lexically hard words than lexically easy words. The first prediction has
209 been borne out experimentally in studies with normal-hearing adults (Sommers & Danielson,
210 1999; Taler et al, 2010; Helfer & Jesse, 2015), children (Eisenberg et al, 2002), cochlear
211 implant users (Kaiser et al, 2003; Bierer et al, 2015) and native and non-native speakers
212 (Bradlow & Pisoni, 1999); the second prediction has also received some experimental
213 support (Sommers & Danielson, 1999; Taler et al, 2010) and will be further tested in the
214 current study.

215 Lexical access can also be affected by the semantic context provided by the words preceding
216 the target: a certain semantic context can markedly increase the likelihood that a given word
217 will occur. It is commonly found that recognition is better for words in semantically
218 meaningful sentences than words in isolation (Miller et al, 1951; Nittrouer & Boothroyd,
219 1990), and for items in sentences with higher as opposed to lower semantic predictability
220 (Bilger et al, 1984). These findings can also be explained in terms of the NAM: as semantic
221 information builds over the course of a sentence, it increases activation levels for contextually
222 consistent words (Sommers & Danielson, 1999).

223 The phenomenon of retrieval-induced forgetting has also been suggested by some researchers
224 (e.g. Anderson, Bjork & Bjork, 2000; Aslan & Bäuml, 2011) as evidence for the role of
225 active inhibition in lexical access (however, see e.g. MacLeod et al (2003) and Williams &
226 Zacks (2001) for alternative interpretations). Retrieval-induced forgetting refers to a situation

227 in which recall for verbal material suffers when related material (e.g. a member of the same
228 category) has earlier been cued and correctly recalled. This suggests that inhibitory processes
229 suppress relevant but uncued material during the initial recall phase, leading to poorer recall
230 for that same material later.

231 **1.2.2 Age-related changes in inhibition and lexical access**

232 The fact that effects of lexical difficulty and semantic context on word recognition vary
233 through the lifespan has been taken as indicating age-related changes in inhibition. For
234 example, the finding that identification of isolated lexically hard words declined with age,
235 while performance for isolated lexically easy words was comparable for younger and older
236 listeners, was interpreted by Sommers (1996) as reflecting an age-related decline in inhibitory
237 control: since competing words from the target's neighborhood have to be suppressed or
238 inhibited for successful word identification, poorer inhibition would reduce the ability to
239 perform the required suppression of competing words and hence result in lower performance
240 for lexically hard words. Results from the audiovisual (AV) domain have been interpreted in
241 a similar vein: the finding that older adults were disproportionately poorer at identifying
242 words with dense audiovisual neighbourhoods was taken as indicating an age-related decline
243 in inhibition (Dey & Sommers, 2015); this hypothesis was supported by the fact that Stroop
244 scores predicted AV word recognition in older, but not younger, adults. Finally, Sommers &
245 Danielson (1999) attribute Pichora-Fuller et al.'s (1995) finding that older listeners benefitted
246 more from the addition of semantic context than younger listeners to higher activation of
247 contextually consistent words amongst older listeners due to increased linguistic experience.

248 However, it is important to note that several studies have failed to show a relationship
249 between inhibitory abilities and SiN perception (Tamati, Gilbert & Pisoni, 2013; Helfer &
250 Freyman, 2014). It is unclear why these discrepancies arose, but one possibility is that these
251 differences were due, at least in part, to the methodological issues described above. Although
252 all of these studies used Stroop tasks to assess inhibition, they differed in the modality of the
253 task used (auditory versus visual), and in the way in which Stroop interference was
254 calculated. In particular, some used traditional SI scores, which as discussed above may be
255 subject to confounds with generalized slowing and/or sensory decline, while others used
256 adjusted scoring systems that may have accounted for slowing, poor colour vision or both. In
257 order to better understand the relationship between inhibition, SI scores and SiN perception,
258 and to investigate how the predictive relationship between SI scores and SiN perception
259 changes depending on whether or not possible confounds in the SI measures have been taken
260 into account, we assessed the predictive value for SiN perception of SI measures derived
261 from an auditory and a visual Stroop task using scoring methods that did or did not account
262 for possible age-related confounds. If the power of Stroop scores to predict SiN perception is
263 based on their ability to measure inhibition, then a purer inhibitory measure free from age-
264 related confounds should improve prediction. However, Stroop scores may primarily measure
265 more general age-related changes, such as generalised slowing and sensory declines. Since
266 generalised slowing will affect performance across a range of tasks, and sensory declines are
267 likely to be shared across the visual and auditory domains (Linderberger & Baltes, 1994), the
268 predictive relationship between Stroop scores and SiN perception may be based more
269 strongly on these age-related changes than on inhibition. If this is the case, then the
270 traditional, unadjusted SI measures should prove more useful in predicting SiN performance.

271 **2 Hypotheses**

272 **2.1 Different scoring systems**

273 H1: Scoring methods can be devised that do or do not take age-related changes in processing
274 speed and sensory decline (i.e. poorer colour vision) into account. If non-inhibitory age-
275 related changes are independent contributors to Stroop scores alongside inhibitory ability
276 (Melara & Algom, 2003), we would expect a low correlation between traditional scores,
277 which do not account for these age-related changes, and the new scores, which do.

278 H2: Stroop scores can be calculated across all trials, or only across trials which are responded
279 to particularly slowly or quickly. We expect the size of the Stroop effect to be larger on
280 average for the slower trials than the faster trials, since a proportional slowing of both longer
281 and shorter RTs leads to a larger differences between the two overall. If it is true that
282 differences in inhibitory ability are more in evidence when participants take longer to respond
283 (Ridderinkhof et al., 2004), then we also expect to see greater variation in individual Stroop
284 effects when examining slower trials as opposed to faster trials.

285 **2.2 Visual versus auditory tasks**

286 H3: The results from the visual and auditory Stroop tasks will be broadly comparable,
287 assuming that a) inhibition is a modality-independent general cognitive ability, b) inhibition
288 influences individual performance to a greater extent than do task-specific demands and c)
289 the two types of task are tapping into the same ability. If this is not the case, this raises
290 questions about the extent to which the two tasks measure the same aspect of cognition.

291 **2.3 Relationship to SiN tasks**

292 H4: Based on previous studies (Sommers & Danielson, 1999; Janse, 2012) we predict larger
293 Stroop interference (SI) scores to be predictive of worse performance on SiN tasks – that is, a
294 negative relationship between SI scores and SiN scores. If SI scores provide a genuine
295 measure of inhibitory ability, then this relationship should be particularly strong when the
296 SiN stimuli demand high levels of inhibition: at lower (less favourable) SNRs, when
297 sentential context is lacking (i.e. when targets are isolated words), when target words have a
298 low word frequency and/or high neighborhood density, or when semantic context does not
299 aid inference (i.e. when targets appear in low-predictability sentences). It is possible that
300 these effects may be particularly pronounced for those with poorer hearing sensitivity (Helfer
301 & Jesse, 2015).

302 H5: If the relationship between SI scores and SiN perception is partially driven by shared
303 sensory decline, we might expect the predictive power of Stroop interference for speech
304 perception to decrease once sensory decline is taken into account. If, on the other hand, it is
305 the inhibition component of the Stroop task that drives the relationship with speech
306 perception, then a purer measure less affected by sensory change might improve the
307 association between the two measures.

308 H6: Based on previous studies suggesting that differences in inhibitory ability are more in
309 evidence when participants take longer to respond (Ridderinkhof et al., 2004), we expect
310 Stroop scores derived from slower trials be better predictors of SiN perception than scores
311 derived from faster trials or averages across all trials.

312 **3 Material and methods**

313 **3.1 Participants**

314 Participants were 50 adults aged over 60 (mean: 69.5 years, SD: 6.4, range = 61-86) with
315 mild hearing loss. A sample size of N = 50 allowed for the detection of a medium-sized effect
316 ($r = 0.35$) at alpha (two-tailed) = 0.05 with a probability of 80%. This was deemed sufficient

317 given that the most closely related previous studies (Janse et al., 2012; Sommers &
318 Danielson, 1999) typically show medium-to-large effect size correlations. Exclusion criteria
319 were hearing aid use and non-native English language status. This study was carried out in
320 accordance with the recommendations of the University of Nottingham's Code of Research
321 Conduct and Research Ethics, with written informed consent from all subjects. All subjects
322 gave written informed consent in accordance with the Declaration of Helsinki. The protocol
323 was approved by the University of Nottingham's School of Psychology Ethics Committee
324 (ref. 464).

325 Visual accuracy was assessed using a Landolt C Chart, and colour vision was tested using the
326 card version of the City University Colour Vision Test. All participants were able to
327 successfully read a full line of optotypes on the Landolt C Chart at a logMAR value of at
328 least 0.3, with the majority (34) able to read a full line at between -0.1 and 0.1 logMAR. Four
329 participants failed the Colour Vision Test, and the same group also verbally reported colour
330 blindness; these participants were excluded from the visual Stroop task. No other participant
331 reported any difficulty in reading the test materials for the visual Stroop task. Two
332 participants were excluded from the auditory Stroop task due to technical failure.
333 Additionally, all participants were screened for mild cognitive impairment (MCI) using the
334 Montreal Cognitive Assessment (MoCA) (mean: 27.86; SD: 1.95).

335 The reported results are part of a larger study into cognitive contributions to speech
336 perception in older adults. Unreported results do not relate to the topics discussed in this
337 paper.

338 **3.2 Auditory measures**

339 Pure-tone air-conduction thresholds (PTA) were collected for nine frequencies between 0.25-
340 8kHz for each ear, following the procedure recommended by the British Society of
341 Audiology (British Society of Audiology, 2011) using an Interacoustics Audiometer AT235
342 (Interacoustics, Middelfart, Denmark) and TDH39P headphones (Telephonics, Farmingdale,
343 NY, USA). Mean thresholds as a function of frequency are presented in Figure 1. As this
344 figure shows, there was considerable variability between participants in terms of hearing
345 acuity, particularly at the higher frequencies.

346 **FIGURE 1 HERE**

347 Speech reception thresholds (SRT) were obtained using 30 sentences from the Adaptive
348 Sentence List (MacLeod & Summerfield, 1990). Sentences were initially presented at 60dB
349 SPL, with a one-down-one-up procedure and step sizes of 10dB down, then 5dB up for the
350 first reversal; the remainder of the trials used a three-down-one-up procedure with a step size
351 of 2dB. The last two reversals were averaged to determine the 79% accuracy point (Levitt,
352 1971). Based on this, all auditory stimuli used throughout the study, including the auditory
353 Stroop stimuli, were presented at 30dB SL – that is, 30 dB above each participant's
354 individual threshold. This procedure was used to partially control for differences in
355 intelligibility in quiet due to the considerable range in participants' hearing sensitivity.

356 **3.3 Stroop tasks**

357 In the visual Stroop task, modelled after Janse (2012), participants were presented with grids
358 formed of 48 boxes in an 8 x 6 arrangement. There were three types of grid: i) a reading grid,
359 consisting of white boxes containing black colour words; ii) a control grid, consisting of
360 coloured boxes containing the string "XXXX" in black; iii) an interference grid, consisting of
361 coloured boxes containing mismatched colour words in black. The colours used were red,

362 blue, green and brown. Using relatively large boxes of colour instead of font colour
363 maximised the opportunity for older participants to clearly see the colours. The distractor
364 words were printed in black and displayed in each box using 20 pt Calibri font. In order to
365 ensure best possible visibility the light in the test room was always at least 880 lux and was
366 set in such a way that each participant could optimally see colours and text without
367 experiencing glare. For i), the task was to read the words aloud as quickly and accurately as
368 possible. For ii) and iii), the task was to name the background colour of the boxes as quickly
369 and accurately as possible. There was a short practice session for each of the 3 tasks.
370 Participants saw two versions of each grid. The total time taken to complete each grid was
371 timed by the experimenter using a stopwatch, and overall scores for each grid type were
372 calculated by averaging the two times obtained.

373 In the auditory Stroop task, modelled after Sommers & Danielson (1999), participants heard
374 two male and two female speakers, and were required to respond as quickly and accurately as
375 possible to the gender of the speaker. Any given trial consisted of one of three words:
376 “mother”, “father” or “person”. These words could therefore be congruent with gender (e.g.
377 female + “mother”), incongruent with gender (e.g. male + “mother”) or neutral (“person”).
378 RTs for gender decisions were obtained via button presses. Participants always used their
379 self-reported dominant hand to respond, and returned their hand to the rest position in front of
380 the button box after the end of each trial. For each trial, the RT was measured from the onset
381 of the sound file; however, the recordings had been trimmed so that, for the words “father”
382 and “person”, voicing started at a similar point in all files (around 13ms after onset for
383 “father”, and around 7ms after onset for “person”). For “mother”, voicing was considered to
384 start early enough that the point of vowel onset was not meaningfully different between any
385 of the four recordings. The location (left/right) of the buttons corresponding to “female” and
386 “male” were swapped for half of the participants. Participants received a short practice
387 session containing all three conditions before the start of the task.

388 **3.4 Speech-in-noise tasks**

389 The SiN tasks varied in both semantic context and lexical difficulty. Semantic context was
390 varied as part of the sentence task, where target words were the final words of low- (LP) and
391 high-predictability (HP) sentences. Stimuli were 112 sentence pairs from a recently
392 developed sentence pairs test (Heinrich et al., 2014). This test, based on the SPIN-R test
393 (Bilger et al., 1984), comprises sentence pairs with identical sentence-final monosyllabic
394 words, which are more or less predictable from the preceding context (e.g. “We’ll never get
395 there at this rate” versus “He’s always had it at this rate”). High and low predictability
396 (HP/LP) sentence pairs were matched for duration, stress pattern, and semantic complexity.
397 Sentences were recorded using a male Standard British English speaker. Only the HP or LP
398 version of a sentence was heard by a single participant.

399 Lexical difficulty was assessed in the word task, where target stimuli were 200 isolated words
400 whose lexical difficulty was varied in terms of word frequency (WF) and neighborhood
401 density (ND). The set of words comprised the 112 final words from the sentence task and an
402 additional 88 monosyllables. WF was measured using the BNC corpus
403 (<http://www.natcorp.ox.ac.uk/>), filtered for nouns (exact form). This corpus was chosen
404 because it both uses British English and also allows particular parts of speech to be isolated:
405 in this case, the measure of interest was the frequency of the target words as nouns, since the
406 sentence contexts led listeners to anticipate a noun target, and as the exact form heard in the
407 sentence, not with potential pluralisations or any other alterations. This limitation was
408 mirrored in the scoring of the SiN task, where only the exact form of a stimulus was scored as
409 correct. ND was determined using N-Watch (Davis, 2005). This tool uses the Celex database

410 to create neighborhood measures using a letter-substitution algorithm, but cross-checks the
411 measures with word frequency to ensure that extremely rare words are not included. This
412 stops over-estimation of ND with respect to most people's vocabulary. It also uses British
413 English. Based on these measures, the 200 words were divided into 4 groups, with WF and
414 ND ranges as follows:

415 **TABLE 1 HERE**

416 All 200 words were re-recorded using a different male Standard British English speaker.

417 All SiN stimuli were presented in speech-modulated noise (SMN). The SMN was created by
418 using an inverse FFT to generate a noise signal with the same long-term average spectrum as
419 the target speech. This noise signal was then modulated in level by dot multiplying it with the
420 absolute value of the smoothed Hilbert transform of the target speech (smoothing was
421 accomplished by convolving the speech envelope with a 46 ms vector of ones). Finally the
422 SMN was scaled to match the RMS level of the target speech. This made the speech signal
423 unintelligible while keeping the long-term average spectrum, level, and temporal envelope of
424 the original signal intact. SiN stimuli were presented in two SNRs to create a more or less
425 adverse listening condition (words at +1dB and -2dB; sentences at -4dB and -7dB). SNR
426 levels were chosen to vary the overall difficulty of the task between 20% and 80% accuracy.
427 Each of the 112 sentence-final words was only heard once by each participant, either in the
428 context of an HP or an LP sentence, and half the sentences of each type were heard with high
429 or low SNR. Each of the 200 words was heard only once, with either high or low SNR, and
430 there were equal numbers of words in each combination of word frequency and neighborhood
431 density categories. After hearing each sentence or word participants repeated as much as they
432 could. Testing was self-paced, and responses were recorded for offline scoring.

433 **3.5 Procedure**

434 Testing was carried out in a double-wall sound-attenuating booth (Industrial Acoustics
435 Company (IAC), Winchester, UK) using Sennheiser HD280 headphones. All testing was in
436 the left ear only. The SiN and Stroop tasks formed part of a larger battery of tests, which were
437 administered over the course of two sessions around a week apart. The two SiN tasks (words
438 and sentences) were always tested in different sessions; the two Stroop tasks (auditory and
439 visual) were tested in different sessions wherever possible, which was the majority of cases.
440 The order of SiN tasks was counterbalanced across participants. There was no systematic
441 pairing of SiN and Stroop tasks within sessions.

442 **3.6 Modelling**

443 In all cases, the outcome measure was speech intelligibility as measured in RAUs
444 (Studebaker, 1985). A number of stimulus-based variables were coded as categorical
445 predictors: semantic predictability (LP/HP) of sentence-final words; word frequency
446 (high/low) and neighborhood density (high/low) of isolated words; speech type
447 (sentences/words) of words and sentences; SNR (high/low). In addition, the following
448 listener variables were coded as continuous predictors: Stroop score (on either the auditory or
449 visual Stroop tasks, using a specified scoring system), and PTA. The PTA variable was
450 calculated by averaging the obtained thresholds at all tested frequencies for each participant,
451 and then centering these values.

452 The relationship between predictor and outcome variables was assessed in a series of linear
453 mixed models (LMMs) using ML estimation, with predictor variables as fixed effects and
454 Type 3 SS. All models included participants as random effects.

455 A backwards stepwise procedure was used to determine the final set of predictors for each
 456 model.¹ This procedure was implemented through manual checking and effect removal. All
 457 analyses were performed in IBM SPSS Statistics 21.

458 **4 Results**

459 **4.1 Mean results for speech-in-noise (SiN) perception**

460 Mean intelligibility values for all SiN conditions are given in Table 2.

461 TABLE 2 HERE

462 Repeated-measures ANOVAs were conducted to investigate group differences in word and
 463 sentence intelligibility due to stimulus-based predictor variables. For intelligibility of
 464 sentence-final words, a semantic predictability (LP/HP) x SNR (low/high) within-subjects
 465 ANOVA showed significant main effects of both predictability ($F(1, 49) = 571.72$; $MSE =$
 466 91.67 , $p < 0.001$, $\eta^2 = 0.921$; $HP > LP$) and SNR ($F(1, 49) = 168.54$; $MSE = 76.81$, $p <$
 467 0.001 , $\eta^2 = 0.775$; $easy > hard$), but no predictability x SNR interaction. For intelligibility of
 468 isolated words, a word frequency (low/high) x neighborhood density (low/high) x SNR
 469 (low/high) within-subject ANOVA showed significant main effects of word frequency (WF)
 470 ($F(1, 49) = 111.67$; $MSE = 37.37$, $p < 0.001$, $\eta^2 = 0.695$; $high > low$), neighborhood density
 471 (ND) ($F(1, 49) = 33.89$; $MSE = 70.11$, $p < 0.001$, $\eta^2 = 0.409$; $low > high$) and SNR ($F(1, 49)$
 472 $= 120.69$; $MSE = 66.54$, $p < 0.001$, $\eta^2 = 0.711$; $easy > hard$); additionally, a significant WF x
 473 ND interaction ($F(1, 49) = 180.40$; $MSE = 54.53$, $p < 0.001$, $\eta^2 = 0.786$) indicated that words
 474 with both a high word frequency and a low neighborhood density were more intelligible than
 475 words in the other three conditions (Bonferroni-corrected at $p = 0.05$).

476 **4.2 Visual Stroop**

477 **4.2.1 Calculating Stroop scores**

478 The Stroop Interference measure (SI) traditionally used in the literature (MacLeod, 1991) is
 479 calculated as follows:

480
$$[1] vSI_{raw} = C_i - C_n$$

481 The mean for C_n was 31.66s (SD = 5.41s); the mean for C_i was 47.13s (SD = 8.14s); and in
 482 all cases the difference between them was positive (i.e. $C_i > C_n$). The mean difference

¹ First, the most complex model was run (i.e. full factorial: all main effects and all possible interactions). Then, non-significant effects were removed one level at a time. For example, if the highest-level interaction was a 4-way interaction and was not significant, it was removed. The model was subsequently re-run. All non-significant 3-way interactions were then removed, and the model was re-run. All non-significant 2-way interactions were then removed, and so on. If a previously significant higher-order interaction lost significance at any stage, this interaction was removed immediately before any further modifications are made. As a general rule, the principle of marginality was observed. As a consequence, if a higher-level interaction was kept in the model, the nested lower-level interactions were also retained. For example, if $A*B*C$ was kept in the model, then the model also included $A*B$, $A*C$ and $B*C$. These relevant nested interactions are called "marginal effects". As this approach has repercussions with regard to model parsimoniousness, a balance between the competing demands of marginality and parsimony was needed. This was achieved by keeping these guidelines in mind: (1) Even if the highest-level interaction was significant, it was not included in the model if it contained 5 or more factors. This allowed the models to be reasonably trimmed in the first instance. (2) A lower-level significant 5- or 4-way interaction was only kept in the model if it contained the Stroop variable. (3) All significant and/or marginal 3-way and 2-way interactions were included, regardless of whether they contained the Stroop variable. (4) All main effects were kept in the model at all times.

483 between RTs in the two conditions for the current dataset was 15.5s (SD = 4.49s) overall,
484 which represents a mean of 0.32s (SD = 0.09s) per item (word).

485 One problem with using the traditional SI measure as an estimate of inhibition in older adults
486 is that there can be age-related changes in general processing speed (Ben-David & Schneider,
487 2009). This would be expected to slow performance on incongruent (Ci) and neutral (Cn)
488 trials by the same factor, leading to different absolute increases – which in turn lead to larger
489 SI values when the difference between the two conditions is calculated. A possible way to
490 account for this age-related change and minimise its effect on interference estimates is to use
491 a normalised measure of Stroop interference. This can be calculated as follows:

$$492 \quad [2] \text{vSI}_{\text{norm}} = C_i/C_n$$

493 In the case of the current dataset this gives a mean score of 1.49 (SD = 0.14).

494 Another problem with the visual SI measure is that the different age-related trajectories for
495 colour vision (declining) and reading speed (stable) mean that colour naming RTs in the
496 neutral condition (Cn) may slow with age relative to reading speed (Rn) (Salthouse & Meinz,
497 1995). The Stroop effect originates from the difference in time course between colour naming
498 in the presence versus absence of a readable distracting colour word. If colour naming slows
499 while word reading remains unchanged with age, then there will be a greater difference in
500 processing speed between the colour naming and reading dimensions, and this puts
501 participants at greater risk of inhibition failure in the incongruent (distractor) condition: that
502 is, if a participant's colour naming speed is relatively slow compared to their reading speed,
503 they have to suppress the irrelevant word for longer, and this increases their chances of
504 experiencing an inhibition failure.

505 Melara & Algom (2003) refer to the discrepancy between access to words and colour names
506 as the Dimensional Imbalance (DI) i.e.

$$507 \quad [3] \text{DI} = C_n - R_n$$

508 Thus a large DI score indicates a slow colour naming speed relative to reading speed. Melara
509 & Algom found DI to be strongly positively correlated with Stroop interference (SI) as
510 measured by [1]: larger DI scores (relatively slow colour naming speeds) were associated
511 with larger Stroop effects.

512 If an increased dimensional imbalance indeed contributes to larger SI (inhibitory failure) in
513 older adults, then it needs to be taken into account when calculating inhibition ability. There
514 are two possible ways to do this. The first is to calculate a standardised Ci using the DI score,
515 as follows:

$$516 \quad [4] \text{vSI}_{\text{standard}} = C_i/\text{DI}$$

517 This factors out the part of Ci which is determined by DI. As a result, differences in colour
518 naming speed relative to reading speed are controlled for, leaving only the portion which
519 represents “true” inhibitory ability.

520 An alternative approach is to use residuals. For a linear regression modelled as $C_{i_j} = \alpha + \beta \text{DI}_j$
521 $+ \epsilon_i$, the residuals can be calculated as:

$$522 \quad [5] \text{vSI}_{\text{res}} = y_{C_i} - \hat{y}_{C_i}$$

523 This method regresses Ci on DI, and then takes the unstandardised residual (i.e. the
524 difference between the observed Ci value (y_{C_i}) and the predicted Ci value (\hat{y}_{C_i})) for each
525 participant. These residuals represent the difference between a participant's observed Ci score

526 relative to what their DI score would predict: a residual near to 0 indicates that the observed
527 C_i score is very similar to what the DI score would predict, suggesting that DI explains
528 almost all of the increase in C_i relative to C_n . A positive residual suggests that the observed
529 C_i score is higher than what could be predicted by DI, indicating “true” inhibitory failure;
530 while a negative residual suggests that the observed C_i is lower than what would be predicted
531 based on DI, and represents “true” inhibitory success. This method thus provides a measure
532 of inhibitory control free from the effects of visual sensory decline. It also accounts for
533 general cognitive slowing since, like [2], it is a relational measure. One issue with this
534 method is that the residual scores depend on the performance of the sample – that is, the
535 predictive relationship between DI and C_i is derived only from the study participants, who
536 may not be representative of the wider population. It would be preferable to independently
537 derive a “gold-standard” relationship between DI and C_i ; however, this has not yet been
538 done, and so for the current study we must rely on the data from our sample alone.

539 **4.2.2 The relationship between visual Stroop scores and speech-in-noise (SiN)** 540 **perception**

541 This section examines the predictive value of visual Stroop interference for SiN perception in
542 high and low predictability sentences and for single words varying in word frequency and
543 neighborhood density. Predictive power for SiN perception was investigated for two
544 measures of visual Stroop interference: vSI_{raw} , the traditional measure for Stroop interference
545 unadjusted for sensory decline, and vSI_{res} , the new measure of Stroop interference that takes
546 general age-related slowing as well as sensory decline into account. The predictive
547 relationship between each of the visual Stroop scores and performance on the sentence task,
548 the word task and the sentence and word tasks combined, are presented in Tables 3, 4 and 5
549 respectively. The analyses combining the scores from the sentence and word tasks (Table 5)
550 was included in order to directly compare the predictive effect of Stroop scores across the
551 two outcome measures. In a second step, PTA was added to each set of analyses in order to
552 examine how it modified the predictive effect of the Stroop scores.

553 Tables 3-5 indicate, for each combination of model type and dataset, a) whether a predictive
554 effect of the Stroop measure on SiN performance was present, and what the nature of the
555 effect was; and b) what, if any, significant interactions between the Stroop measure and
556 stimulus-based variables or PTA were present. The effects are described as rate of change
557 where a positive slope indicates an average increase in SiN performance with every
558 additional increase in Stroop interference, while a negative slope indicates an average
559 decrease in SiN performance with every additional increase in Stroop interference. Based on
560 our hypotheses, we expect negative slopes. While PTA was always entered as a continuous
561 predictor, we use a categorical median split when reporting and discussing its effects, because
562 it allows for clearer descriptions, particularly of complex interactions. The tables do not list
563 significant interactions if they do not involve the Stroop measure. The AIC value is included
564 for each model as an indication of goodness-of-fit, with lower AIC values corresponding to a
565 better fit.

566 TABLE 3 HERE

567 TABLE 4 HERE

568 TABLE 5 HERE

569 The models reveal a complex pattern of results with the direction of the relationship between
570 the vSI measures and SiN performance, as well as the strength of the relationship, depending
571 on the scoring method and characteristics of the stimulus and the listener. However, in all

572 cases, the inclusion of PTA in the model enhanced model fit (i.e. produced a lower AIC
573 value).

574 We will now examine, for each dataset in turn, how the nature of the relationship between
575 Stroop scores and SiN performance was modulated by stimulus-based variables and PTA for
576 each Stroop scoring method.

577 4.2.2.1 Sentence perception

578 *Traditional (vSI_{raw}) measure.* There was no predictive effect of the Stroop measure
579 overall, and stimulus-based predictors did not modulate the predictive effect of Stroop
580 interference. There was also no modulating effect of PTA.

581 *Adjusted (vSI_{res}) measure.* While there was no predictive main effect of Stroop
582 interference, an interaction of vSI_{res} x Pred x SNR indicates that the predicted negative
583 relationship between Stroop scores and sentence perception was seen for the high
584 predictability (HP) sentences in the harder SNR, and for the low predictability (LP) sentences
585 in the easier SNR, but not for the HP sentences in the easier SNR or the LP sentences in the
586 harder SNR. There was no modulating effect of PTA.

587 4.2.2.2 Word perception

588 *Traditional (vSI_{raw}) measure.* While there was no predictive main effect of Stroop
589 interference, an interaction with neighborhood density (ND) indicates that the observed
590 relationship between vSI_{raw} and word perception was more negative for words with less dense
591 neighborhoods. Once PTA was added to the model, an interaction of vSI_{raw} x ND emerged.
592 This indicates that the predictive effect of Stroop scores was strongest for low ND words.
593 This interaction was modulated by SNR and PTA, indicating that the relationship between
594 Stroop scores and SiN perception changed in different ways across ND and SNR conditions
595 for listeners with better and worse hearing. Specifically, the relationship was negative for
596 those with PTA, but was more mixed for those with good PTA, being positive for high ND
597 words in the easier SNR and approaching zero for both ND conditions in the harder SNR

598 *Adjusted (vSI_{res}) measure.* There was no main effect of Stroop interference and no
599 modulating effects of stimulus-based variables. The interaction of vSI_{res} x ND indicates that
600 the predictive effect of Stroop scores was strongest for low ND words. This interaction was
601 further modulated by PTA, indicating that the relationship between Stroop scores and SiN
602 perception changed in different ways for the two ND conditions when examining listeners
603 with better and worse hearing. Specifically, for those with worse hearing the Stroop/SiN
604 relationship was more negative for low ND words but less negative for high ND words when
605 compared to those with better hearing

606 4.2.3.3 Speech (combined dataset)

607 *Traditional (vSI_{raw}) measure.* There was no predictive main effect of Stroop measure.
608 An interaction with Type indicates that the predictive effect of Stroop scores for SiN
609 perception differed in direction between sentences and words, being negative for the word
610 task and positive for the sentence task. PTA did not modulate the found relationships.

611 *Adjusted (vSI_{res}) measure.* There was no main effect of Stroop interference and no
612 modulating effects of stimulus-based variables, and no modulating effect of PTA

613 In summary, the predictive effect of visual Stroop scores for SiN perception is similar
614 in some respects regardless of the scoring method. Both scoring systems reveal influences of
615 lexical factors (sentence predictability and word neighborhood density), and neither system

616 shows a large effect of PTA. However, there are also important differences between the two
 617 scoring systems. In particular, the direction of the Stroop/SiN relationship changes depending
 618 on the type of target speech when using the traditional scoring method, whereas it is
 619 consistently negative across speech types for the vSI_{res} method.

620 4.3 Auditory Stroop (all trials)

621 4.3.1 Calculating Stroop scores

622 The auditory Stroop task resulted in three measures for each participant: average RT for
 623 neutral trials (aRT_n), congruent trials (aRT_c) and incongruent trials (aRT_i). Initial inspection
 624 of the data revealed that not all four speakers produced Stroop interference effects for every
 625 participant. We therefore analysed for each participant the responses to the female and male
 626 speaker who produced, for that participant, the largest overall traditional Stroop interference
 627 (RT_i – RT_n). Speakers M1 and M2 were chosen 13 and 35 times respectively, speakers F1
 628 and F2 25 and 23 times respectively. Following Green & Barber (1981), only correct trials
 629 (from the aRT_i and aRT_n conditions were included in any analysis.

630 Congruent trials are usually included in Auditory Stroop tasks, and previous studies (Green &
 631 Barber, 1981; Jerger et al, 1988) have found a facilitation effect (i.e. faster responses to
 632 congruent than neutral trials), although this is not always the case (Sommers & Danielson,
 633 1999). Using a 1-way repeated-measures ANOVA (Greenhouse-Geisser corrected for
 634 violations of sphericity) with aRT_n, aRT_c and aRT_i as within-subject levels of condition, we
 635 found a main effect of condition ($F(2, 79) = 53.40$; $MSE = 0.005$, $p < 0.001$, $\eta^2 = 0.532$).
 636 Post-hoc testing showed an interference effect but no facilitation effect (aRT_i > aRT_c, aRT_i >
 637 aRT_n, aRT_c = aRT_n (Bonferroni-corrected at $p = 0.05$)).

638 The traditional Stroop Interference measure (SI) for the auditory Stroop is calculated
 639 analogously to the visual Stroop:

$$640 \quad [6] \text{ aSI}_{\text{raw}} = \text{aRT}_i - \text{aRT}_n$$

641 The mean aRT_i (per item) is 1.33s (SD = 0.23s), the mean aRT_n is 1.20s (SD = 0.21s), and
 642 aRT_i is higher than RT_n for all but 3 listeners. The mean difference between RTs in the two
 643 conditions for the current dataset is 0.13s (SD = 0.09s) per item (word). This difference is
 644 smaller than for the visual Stroop.

645 As explained above, the issue of generalised slowing makes the traditional Stroop (SI)
 646 measure problematic: if aRT_i and aRT_n increase by the same factor, SI will also increase; this
 647 means that a larger SI may reflect slowing rather than paucity of inhibition. Normalised SI
 648 was proposed as one means of addressing the issue of generalised slowing, and can be
 649 calculated for the auditory Stroop as follows:

$$650 \quad [7] \text{ aSI}_{\text{norm}} = \text{aRT}_i / \text{aRT}_n$$

651 In the case of the current dataset this gives a mean score of 1.11 (SD = 0.08).

652 4.3.2 The relationship between auditory Stroop scores and speech-in-noise (SiN) 653 perception

654 This section examines the predictive value of auditory Stroop interference for SiN perception
 655 in high and low predictability sentences, and for single words varying in word frequency and
 656 neighborhood density. As before, performance in these conditions was predicted by one of two
 657 auditory Stroop interference measures: aSI_{raw}, the traditional measure for Stroop interference,
 658 or aSI_{norm}, a measure of Stroop interference that takes generalised slowing into account. The

659 relationship between each Stroop measure and SiN perception, as characterised by a series of
660 LMMs, is summarised in Tables 6 to 8. In all cases, the first part of the table presents the results
661 when Stroop interference and stimulus-based variables are the only predictors of SiN
662 performance. The second part of each table presents the results when PTA is considered in
663 addition to Stroop interference and stimulus-based variables.

664 TABLE 6 HERE

665 TABLE 7 HERE

666 TABLE 8 HERE

667 For both auditory Stroop scoring systems, the overall relationship between Stroop scores and
668 SiN perception is mostly positive. This is truer for the normalised (aSI_{norm}) scores than the
669 traditional (aSI_{raw}) scores, since Stroop scores never reach significance as a main effect when
670 using the aSI_{raw} scoring method, but is significant across all datasets when using the aSI_{norm}
671 measure without PTA. As before, including PTA improved the fit of the model in all cases.

672 We will now examine, for each dataset in turn, how the nature of the relationship between
673 Stroop scores and SiN performance was modulated by stimulus-based variables and PTA for
674 each Stroop scoring method.

675 4.3.2.1 Sentence perception

676 *Traditional (aSI_{raw}) measure.* There was no main effect of Stroop interference and no
677 modulating effects of stimulus-based variables or PTA.

678 *Adjusted (aSI_{norm}) measure.* There was a positive predictive main effect of Stroop
679 scores but no modulating effects of stimulus-based variables on their own. When PTA was
680 added as an additional predictor an interaction of $aSI_{norm} \times \text{Pred} \times \text{SNR} \times \text{PTA}$ emerged,
681 which indicates that the predictive strength, but not the direction, of Stroop interference for
682 speech perception in a particular condition depended on a person's hearing ability.

683 4.3.2.2 Word perception

684 *Traditional (aSI_{raw}) measure.* While there was no predictive main effect, an
685 interaction of $aSI \times \text{SNR}$ indicates that the positive predictive effect of Stroop scores on SiN
686 performance was stronger at the harder SNR. There was also no modulating effect of PTA

687 *Adjusted (aSI_{norm}) measure.* As for aSI_{raw} above.

688 4.3.2.3 Speech (combined dataset)

689 *Traditional (aSI_{raw}) measure.* Again, there was no predictive main effect of Stroop,
690 but an interaction with SNR indicating a stronger positive predictive effect at the
691 harder SNR. There was no modulating effect of PTA

692 *Adjusted (aSI_{norm}) measure.* As for aSI_{raw} above.

693 In summary, the predictive relationship between auditory Stroop scores and SiN perception is
694 in some ways similar for auditory Stroop scores calculated using the traditional method
695 (aSI_{raw}) and the normalisation method (aSI_{norm}). For both scoring methods, the Stroop/SiN
696 relationship is positive overall and stronger at the more challenging SNR. However, there are
697 also important differences. In particular, the traditional Stroop scores (aSI_{raw}) have no
698 predictive value for performance on the sentence task, whereas the aSI_{norm} scores do.

699 4.4 Auditory Stroop (slow vs. fast trials)

700 As discussed in the introduction, using average measures across all trials of a Stroop task may
701 not be the most efficient way of quantifying inhibition and its failure. We know that
702 inhibition takes time to build up, and that its effects may therefore be strongest for each
703 participant's slowest RTs for incongruent trials (Ridderinkhof, 2002; Ridderinkhof et al,
704 2004; Roelofs et al, 2011). During these trials the distractor has the greatest chance to
705 interfere, but inhibition also has the greatest potential to be deployed by those who can
706 successfully do so; thus individual differences in inhibitory abilities will be most in evidence,
707 since the disparity between those able to successfully deploy inhibition and those less able to
708 do so will be largest during these trials (Roelofs et al, 2011). To assess this, slow and fast
709 trials must be analysed separately. This type of differential analysis of single trials is usually
710 done using delta plots and delta scores.

711 Delta scores were calculated using neutral (aRT_n) and incongruent (aRT_i) conditions. For
712 each participant and each condition, the trials were sorted by RT, and then split into equally-
713 sized quintiles. The average RT was calculated for each quintile in each condition. Mean RT
714 per quintile is the averaged RT across aRT_n and aRT_i for a given quintile. Delta RT per
715 quintile is calculated as mean aRT_i minus mean aRT_n for a given quintile. When averaged
716 over all participants the grand mean RT and grand delta RT can be obtained for each quintile.
717 It is worth noting that, since delta RT per quintile is obtained by calculating $aRT_i - aRT_n$ for
718 that quintile, it is conceptually no different to using the traditional (aSI_{raw}) measure (see
719 equation [6] above). It is the same calculation, but performed using only a subset of trials.

720 Delta plots show grand mean RTs plotted against grand delta RTs for the five RT quintiles
721 (Q1-Q5). Since the delta RT measure compares conditions with and without distractors, and
722 interference from distractors increases over time, the plots typically show an overall increase
723 in delta RTs as mean RTs increase. Individual differences in the build-up of inhibition are
724 expressed in a delta plot by differences in this relationship between mean and delta RTs
725 (Ridderinkhof et al, 2004). Those who are not successfully inhibiting show a monotonic
726 increase in delta RT as mean RT increases. In contrast, those who are successfully engaging
727 inhibition initially show a monotonic increase in delta RT, but for the slowest trials the
728 relationship between delta RT and mean RT will become less steep, flatten out or even
729 become negative. Delta plots therefore allow us to focus on those trials that both allow and
730 require the most inhibition for successful performance, thereby maximizing the chance of
731 seeing individual differences in inhibitory ability.

732 Because participants varied widely in overall RTs, we divided each delta RT by its relevant
733 mean RT to get a normalised delta score, called hereafter aSI_{ndelta} . These scores are plotted in
734 Figure 2.

735

FIGURE 2 HERE

736 A repeated-measures 1-way ANOVA with quintiles as within-subject effects (Greenhouse-
737 Geisser corrected for violations of sphericity) showed a main effect of quintile ($F(2,84) =$
738 18.69 , $MSE = 0.007$; $p < 0.001$, $\eta^2 = 0.284$), and subsequent pairwise comparisons
739 (Bonferroni corrected at $p = 0.05$) revealed that Q5 had significantly higher normalised delta
740 scores compared to all other quintiles, which were not significantly different from each other.
741 However, as Figure 2 shows, Q5 produced not only the largest delta scores (largest Stroop
742 effects) on average, but also the largest variation in scores: the standard deviation of scores in
743 Q5 is 0.12s, compared to a range of 0.05-0.07 for Q1-4. This is in concordance with
744 Ridderinkhof et al. (2004), and also suggests that Q5 is most likely to reveal differential
745 associations between the auditory Stroop measure and SiN perception. If Ridderinkhof and
746 colleagues are correct that there is not enough time for inhibition to become sufficiently

747 strong and/or be successfully deployed during participants' fastest responses, then Q1 should
748 not only show smaller Stroop effects on average and a limited variation in scores, as
749 demonstrated above, but should also have only limited predictive value for performance on
750 the SiN perception tasks.

751 To summarise: delta scores can be used to examine Stroop interference (SI) in different
752 subsets of trials from a Stroop task. Conceptually, these delta scores are the same as the
753 traditional (aSI_{raw}) measure, but calculated using only those trials which fall in a given
754 section of a participant's RT distribution. We are interested in assessing SI derived from the
755 slowest quintile (Q5) and fastest quintile (Q1) of each participant's trials. The slowest trials
756 are used because individual differences in performance on inhibition tasks have been shown
757 to be greatest in this quintile (Ridderinkhof et al, 2002; Ridderinkhof et al, 2004), thus giving
758 us better statistical power to observe links with SiN perception. This larger variation in
759 individual differences is hypothesised to be due to slow trials better revealing individual
760 differences in inhibition (Ridderinkhof et al, 2004; Roelofs et al, 2011). For this reason, we
761 hypothesise that delta scores from Q5 will correlate more strongly with SiN perception than
762 scores from Q1: that is, if SiN perception is determined, at least in part, by inhibitory ability,
763 then SiN scores should correlate more strongly with measures which better reveal differences
764 in inhibitory ability.

765 **4.4.1 The relationship between auditory Stroop delta scores and speech-in-noise (SiN)** 766 **perception**

767 This section examines the predictive value of the two auditory Stroop delta score measures
768 for SiN perception in the six SiN conditions. Two auditory Stroop interference measures were
769 used: $aSI_{ndeltaQ5}$ as a measure of interference derived from the slowest trials; and $aSI_{ndeltaQ1}$ a
770 measure of interference derived from the fastest trials. The relationship between each of these
771 measures and SiN perception, as characterised by a series of LMMs, is summarised in Tables
772 9 to 11.

773 TABLE 9 HERE

774 TABLE 10 HERE

775 TABLE 11 HERE

776 We will now examine, for each dataset in turn, how the nature of the relationship between
777 Stroop scores and SiN performance was modulated by stimulus-based variables and PTA for
778 each Stroop scoring method.

779 **4.4.1.1 Sentence perception**

780 *Slowest ($aSI_{ndeltaQ5}$) trials.* While there was no predictive main effect of the Stroop
781 measure, an interaction of $aSI_{ndeltaQ5}$ x Pred x SNR indicates that the positive slope predicting
782 SiN performance from Stroop interference was steeper for high predictability (HP) sentences
783 in the more challenging SNR, and for low predictability (LP) sentences in the easier SNR.
784 There was no additional modulating effect of PTA

785 *Fastest ($aSI_{ndeltaQ1}$) trials.* There was no main effect of Stroop interference and no
786 modulating effects of stimulus-based variables or PTA.

787 **4.4.1.2 Word perception**

788 *Slowest ($aSI_{ndeltaQ5}$) trials.* In addition to a positive predictive main effect of Stroop
789 scores, an interaction of $aSI_{ndeltaQ5}$ x SNR indicates that the positive slope predicting SiN

790 performance from Stroop interference was steeper at the harder SNR. This interaction was
791 not modulated by PTA.

792 *Fastest (aSI_{ndeltaQ1}) trials.* There was no main effect of Stroop interference and no
793 modulating effects of stimulus-based variables or PTA.

794 4.4.1.3 Speech (combined dataset)

795 *Slowest (aSI_{ndeltaQ5}) trials.* There was no predictive main effect of Stroop. An
796 interaction with Type indicates that there was a stronger positive predictive effect of Stroop
797 scores for SiN perception in the word than the sentence task. An interaction with SNR
798 indicates that the positive predictive effect of Stroop scores on SiN performance was stronger
799 at the harder SNR. The interaction with Type was modulated by PTA, indicating that the
800 Stroop/SiN relationship varied in strength across SiN type and levels of hearing loss, but
801 remained positive throughout.

802 *Fastest (aSI_{ndeltaQ1}) trials.* There was no main effect of Stroop interference and no
803 modulating effects of stimulus-based variables. An interaction of aSI_{ndeltaQ1} x Type x PTA
804 indicated that the relationship between Stroop scores and SiN perception varied in strength
805 across SiN type and levels of hearing ability, but remained negative overall

806 In summary, the relationship between auditory Stroop scores and SiN perception varies
807 considerably depending on whether the auditory Stroop scores are calculated using either
808 only the slowest responses (aSI_{ndeltaQ5}) or only the fastest responses (aSI_{ndeltaQ1}). First, for
809 aSI_{ndeltaQ5}, the Stroop/SiN relationship is positive overall, stronger for words than sentences
810 for those with poor hearing, and stronger at the more challenging SNR. This stands in
811 contrast to the aSI_{ndeltaQ1} scores, for which the Stroop/SiN relationship is negative overall,
812 stronger for sentences than words for those with poor hearing, and unaffected by SNR.
813 Second, the aSI_{ndeltaQ1} scores have no predictive value for performance on the sentence task,
814 whereas the aSI_{ndeltaQ5} are significantly related to sentence perception. Finally, it is worth
815 noting that the aSI_{ndeltaQ1} scoring method reveals a mixture of positive and negative
816 Stroop/SiN relationships. However, for aSI_{ndeltaQ5} – the scoring method which uses only the
817 very slowest trials – the relationship between Stroop scores and SiN perception is almost
818 always positive.

819 4.5 Intercorrelations of Stroop scoring systems

820 TABLE 12 HERE

821 Table 12 shows the intercorrelations of all six Stroop scoring systems used in the
822 current study. The scores for the two visual Stroop scoring methods, vSI_{raw} and vSI_{res}, are
823 highly positively correlated. The scores for the two auditory Stroop scoring methods which
824 use data from all trials, aSI_{raw} and aSI_{norm}, are also highly correlated. The auditory Stroop
825 scores which use data from all trials are also highly correlated with the auditory Stroop score
826 derived from the slowest trials (aSI_{ndeltaQ5}), and moderately correlated with the auditory
827 Stroop scores derived from the fastest trials (aSI_{ndeltaQ1}). However, the scores from the slow
828 and fast trials (aSI_{ndeltaQ5} and aSI_{ndeltaQ1}) are not correlated with each other. There are no
829 significant correlations between the scores from either of the visual Stroop scoring systems
830 and any of the scores from the auditory Stroop scoring systems.

831 5 Discussion

832 Inhibition is a key cognitive ability, and has been suggested to be important for speech-in-
833 noise perception. However, existing attempts to connect inhibitory abilities to performance

834 on speech-in-noise tasks may have been complicated by methodological issues regarding the
835 use of Stroop tasks. One widely-used method for measuring inhibition is the colour-word
836 Stroop task (Stroop, 1935), which uses visual stimuli and exploits the difference in
837 processing time between reading and colour naming. More recently, auditory Stroop tasks
838 have been developed (Green & Barber, 1981; Morgan & Brandt, 1989) that are designed to
839 measure auditory inhibitory abilities. However, the relationship of these two types of Stroop
840 task, and the question of whether or not they assess the same underlying ability, is not clear.
841 Another issue concerning all Stroop tasks is the question of which scoring system is the most
842 appropriate for estimating inhibitory ability independent of sensory contributions. This
843 question is particularly pertinent to research involving older adults, where it is important not
844 to misattribute sensory changes to changes in cognition. Here we set out to investigate both of
845 these questions – that is, whether auditory and visual Stroop tasks assess similar aspects of an
846 underlying concept, and how the use of different scoring systems that either do or do not take
847 sensory changes into account affects the results. In all cases, the outcome of interest was the
848 way in which a particular Stroop task analysed using a particular scoring method related to
849 and predicted performance on a set of speech-in-noise tasks.

850 We used two Stroop tasks, a visual and an auditory. For the visual Stroop task we explored
851 two scoring methods: the traditional Stroop Interference measure (vSI_{raw}), and a residuals-
852 based measure designed to account for both generalised slowing and declines in colour vision
853 (vSI_{res}). For the auditory Stroop data, we explored four scoring methods: the traditional
854 Stroop Interference measure (aSI_{raw}), a normalised version of the traditional measure
855 designed to account for generalised slowing (aSI_{norm}), a normalised measure of interference
856 for each participant's slowest trials ($aSI_{ndeltaQ5}$) and a normalised measure of interference for
857 each participant's fastest trials ($aSI_{ndeltaQ1}$).

858 The speech tasks were selected to probe various ways in which inhibition could be important
859 for speech perception. First, all target speech was presented in noise because it has been
860 suggested that good inhibition is needed to reduce the susceptibility to background noise
861 (Janse, 2012). Second, target speech was varied in either a) word frequency and
862 neighborhood density for single words or b) semantic context for sentences, because these
863 lexical and semantic characteristics have been hypothesized to tax inhibition to different
864 extents (Sommers & Danielson, 1999).

865 **5.1 Different scoring systems**

866 *H1: If age-related changes in processing speed and sensory decline are independent*
867 *contributors to Stroop scores in addition to inhibitory ability (Melara & Algom, 2003), we*
868 *expect a low correlation between traditional scores (vSI_{raw}), which do not take them into*
869 *account, and the new scores (vSI_{res}), which do.*

870 This hypothesis was assessed using the visual Stroop data. As shown in Table 12, correlations
871 are extremely high between the vSI_{raw} and vSI_{res} measures. This suggests one of two possible
872 interpretations: first, that the participants in this study had not experienced significant
873 declines in colour vision; or alternatively, that sensory decline and inhibitory ability are not
874 independent processes. The first interpretation is unlikely given Ben-David and colleagues'
875 (2009) meta-analysis, which strongly suggests that sensory decline amongst older people is
876 widespread. The second interpretation implies that the two processes deteriorate in a
877 comparable fashion, so that scores which account for sensory decline will nevertheless
878 decline at a similar rate to those which do not. We think that this is a more likely explanation
879 of our data.

880 *H2: We expect to see larger Stroop interference overall and greater variation in individual*
881 *Stroop scores when examining slower trials.*

882 We investigated this hypothesis using the auditory Stroop data. Both of these hypotheses
883 were supported by the data. To examine the slowest and fastest trials, we used normalised
884 delta scores per quintile – that is, for each quintile of the RT distribution, we calculated
885 Stroop interference effects and then normalised them according to the mean RT of the
886 incongruent and neutral trials under examination. Despite using scores that were adjusted for
887 overall RT, we nevertheless found the largest Stroop effects overall for Q5 – the very slowest
888 trials. We also found the widest range of Stroop scores in Q5, which implied that Stroop
889 effects were not uniformly large in this quintile, but instead varied from only marginally
890 higher than those in faster quintiles to substantially increased. This supports the proposal of
891 Ridderinkhof and colleagues (2014) that, although slower RTs allow for greater interference
892 from a distractor, they also allow inhibition to build up and be deployed and, as a result, it is
893 during these slowest responses that inhibitory differences become most apparent.

894 **5.2 Visual versus auditory tasks**

895 *H3: If inhibition is a modality-independent general cognitive ability, and if it influences*
896 *individual performance to a greater extent than do task-specific demands, then the results*
897 *from the visual and auditory Stroop tasks should be broadly comparable.*

898 Table 12 shows that the visual Stroop measures were entirely uncorrelated with the auditory
899 Stroop measures; furthermore, the only correlation which neared significance – that of vSI_{raw}
900 with $aSI_{\Delta Q5}$ – was negative, meaning that the two measures in fact showed opposite trends.
901 This was in stark contrast to within-task correlations, which showed that the two visual
902 Stroop scoring systems were closely correlated with each other, and the auditory Stroop
903 scoring systems were also closely correlated. The only exception to this was the correlation
904 between the $aSI_{\Delta Q1}$ and $aSI_{\Delta Q5}$ measures, which was only moderate. This finding raises
905 questions about the extent to which the two tasks measure the same aspect of cognition, either
906 because separate inhibitory functions operate in different modalities and/or because task-
907 specific demands outweighed the influence of inhibitory abilities in determining individual
908 differences.

909 **5.3 Relationship to SiN tasks**

910 *H4: Larger Stroop interference scores are expected to be predictive of worse performance on*
911 *SiN tasks, particularly when the SiN stimuli demand high levels of inhibition i.e. in less*
912 *favourable SNRs, for isolated targets words, target words with low word frequency and/or*
913 *high neighborhood density, or for low-predictability sentential context. These effects may be*
914 *more pronounced for listeners with poorer hearing.*

915 We predicted a negative Stroop/SiN relationship, with larger Stroop effects predicting lower
916 scores (i.e. worse performance) on SiN tasks. However, we only found this negative
917 relationship in certain SiN conditions, and for certain listeners. For the auditory Stroop task,
918 the overall direction of the relationship to SiN perception changed depending on which
919 section of the RT distribution was under examination: for scores derived from the very
920 slowest responses ($aSI_{\Delta Q5}$), the relationship was almost always positive; for scores derived
921 from the very fastest responses ($aSI_{\Delta Q1}$), the relationship was generally negative – but even
922 using these scores, some stimulus types, in conjunction with listener characteristics, produced
923 a positive Stroop/SiN relationship. The fact that we found a negative Stroop/SiN relationship
924 overall only when using the $aSI_{\Delta Q1}$ scores suggests that participants were engaged in two
925 qualitatively different response modes: that for fast responses and that for slow responses,

926 only the former of which was related to SiN perception in the predicted fashion. The reasons
927 for this are unclear, but it is possible that participants were not always responding as fast as
928 they could, despite instructions to do so. Delaying responses beyond the point at which the
929 correct answer is accessed – for example, to mentally check the response, or because a
930 regular rhythm of responding has been established – may distort Stroop effects in several
931 ways. First, it may make it hard to distinguish between incongruent trials with failed
932 (relatively longer SI) or successful (relatively shorter SI) inhibition, because responses to
933 both are slow; second, if participants delay their responses to trials in the congruent
934 condition, it may make Stroop effects appear smaller than they are, since it becomes harder to
935 distinguish between trials with and without distractors. Distorted Stroop results are less likely
936 to have a meaningful or interpretable relationship to SiN perception. In the case of the current
937 data, if the fastest trials represent “true”, non-delayed responses while the slowest trials
938 represent responses with an artificial delay, this may explain why the predicted Stroop/SiN
939 relationship was seen only for the faster trials.

940 Assuming that the Stroop scores reliably reflected inhibitory abilities, we also expected the
941 (negative) predictive effect of Stroop scores for SiN perception to interact with stimulus
942 parameters such that a stronger effect was seen for those parameter levels which make
943 listening harder and demand higher levels of inhibition. Specifically, these were the harder
944 (as opposed to easier) SNR, isolated words as targets (as opposed to targets presented in
945 sentences), low (as opposed to high) frequency and/or high (as opposed to low) neighborhood
946 density targets, and/or targets in low (as opposed to high) predictability sentences. In some
947 cases, we found this prediction to be true. For example, when using the vSI_{res} method, we
948 found a stronger relationship between Stroop scores and word perception for high
949 neighborhood density words than low neighborhood density words for those with poorer
950 hearing abilities. However, the results are sometimes hard to interpret: for example, we find
951 for many of the auditory Stroop scoring systems that the Stroop/SiN relationship is stronger
952 at the less favourable SNR, and for two of these scoring systems the relationship is also
953 stronger for words as opposed to sentences – but in these cases, the relationship is in the
954 unexpected positive direction, and therefore does not indicate a greater predictive value in the
955 expected sense. Finally, there are also cases in which the results run directly against our
956 hypothesis: for the vSI_{raw} scoring system, we find a stronger negative predictive Stroop/SiN
957 relationship for words with low neighborhood densities, despite the fact that these words
958 should theoretically demand a lower level of inhibition than their high neighborhood density
959 counterparts. Similarly, when using the $aSI_{ndeltaQ5}$ and $aSI_{ndeltaQ1}$ scoring systems we find, for
960 certain listeners (good PTA and poor PTA respectively), a stronger negative predictive
961 Stroop/SiN relationship for sentences as opposed to words, despite the fact that isolated
962 words should tax inhibition more than words presented within a sentential context. These
963 results therefore suggest that, although the sentential context provides additional cues
964 compared to the isolated words, these cues are not working in a consistent fashion to
965 modulate the relationship between Stroop scores and SiN performance. Consequently, the
966 questions of whether or not the Stroop scores genuinely provide a measure of inhibitory
967 abilities, and whether inhibition is involved in SiN perception in a consistent manner, remain
968 unanswered.

969 The suggestion that any effects might be particularly pronounced for those with poorer PTA
970 scores was not generally borne out. There was a very limited role for PTA in the relationship
971 between visual Stroop scores and SiN perception; this is perhaps to be expected given the
972 non-auditory nature of the visual Stroop task. However, PTA played a similarly limited role
973 when looking at the relationship between auditory Stroop scores and SiN perception;
974 furthermore, the nature of those modulating effects which are present is unclear. The

975 somewhat limited role of PTA in the results despite a large range of hearing sensitivity in the
976 tested sample might be explained by the fact that stimuli were presented at 30dB above each
977 listener's individual SRT, which we hoped would to some extent mitigate difficulties caused
978 by poorer hearing abilities.

979 *H5: If correlations between Stroop scores and SiN perception are driven by shared sensory*
980 *decline, we expect the predictive power of Stroop interference for speech perception to*
981 *decrease once sensory decline is accounted for. If the inhibition component drives the*
982 *relationship, then a purer measure might improve the association between the two measures.*

983 For the visual Stroop task, the vSI_{raw} score appears to have slightly greater predictive value
984 for SiN perception than the adjusted vSI_{res} score. As can be seen in Tables 3-5 above, models
985 using the vSI_{raw} score almost always produce smaller AIC values (i.e. a better fit) than
986 models using the vSI_{res} score. These differences are small, with AIC values for models using
987 vSI_{raw} scores being only 1.74 smaller on average; however, this nevertheless suggests that the
988 relationships between visual Stroop scores and SiN perception may rely in part on shared
989 sensory decline. Without a measure of visual sensory decline, this hypothesis cannot be
990 directly tested. At the very least, however, our findings suggest that taking sensory decline
991 into account does not substantially enhance the predictive power of visual Stroop scores for
992 modelling SiN perception.

993 *H6: If Stroop scores derived from slower trials are better able to reveal individual*
994 *differences in inhibitory ability, then these might be better predictors of SiN perception than*
995 *average scores.*

996 For the auditory Stroop task, there was no evidence to suggest that the $aSI_{ndeltaQ5}$ scoring
997 system had greater predictive power for SiN perception than the other methods used. Indeed,
998 as Tables 6-11 show, models using the $aSI_{ndeltaQ5}$ scoring method consistently produced
999 substantially larger AIC values (i.e. a poorer fit) than models using either the aSI_{raw} or aSI_{norm}
1000 methods. The average difference in AIC values between models using the $aSI_{ndeltaQ5}$ scoring
1001 method and those using the aSI_{raw} and aSI_{norm} scores was 35.98 and 39.62 respectively.

1002 **6 Conclusion**

1003 In this study we compared results from several different scoring systems for both visual and
1004 auditory Stroop tasks, and assessed their predictive value with respect to speech-in-noise
1005 perception. The results suggest that these two types of Stroop task may actually be measuring
1006 different aspects of cognition, rather than tapping a single modality-independent general
1007 cognitive ability. The use of different scoring systems changed the relationship of Stroop
1008 scores to speech-in-noise perception. On the one hand, this suggests that different scoring
1009 systems may allow different aspects of participants' responses to be selectively used in
1010 analysis – for example, isolating slower trials to measure the strongest inhibitory effects.
1011 However, it also suggests that traditional Stroop scores may not be reliable measures of
1012 inhibition, but may instead confound inhibitory abilities – or at least those abilities recruited
1013 in speech-in-noise perception –with task-specific demands and participant variables such as
1014 general response speed and visual acuity. Thus caution must be exercised in the use of Stroop
1015 tasks and, if one is used, the scoring system must be carefully selected, particularly if there is
1016 any reason to suspect that participants may be experiencing age-related sensory declines or
1017 generalised slowing. Finally, hearing loss affected the relationship between Stroop scores and
1018 speech-in-noise perception, highlighting the importance of accounting for individual
1019 differences in both demographic factors and sensory acuity when analysing cognitive data.
1020 Indeed, when choosing a cognitive task and/or scoring system, researchers may want to

1021 consider not just the nature of their outcome variable but also the degree to which they wish
1022 to minimise or emphasis the effects of listener variables.

1023 **7 Limitations and further directions**

1024 It must be noted that there are range of cognitive functions referred to as “inhibition”. For
1025 example, Friedman & Miyake (2004) describe three inhibition-related functions:

1026

1027 1) Prepotent Response Inhibition (the ability to deliberately suppress a prepotent response, as
1028 tested in Stroop tasks)

1029 2) Resistance to Distractor Interference (the ability to resist interference from irrelevant
1030 information in the external environment, as tested in e.g. flanker tasks)

1031 3) Resistance to Proactive Interference (the ability to resist intrusions from memory of
1032 information that was previously task-relevant but is now irrelevant)

1033

1034 Using a variety of tasks to assess each function, they found that 1) and 2) were closely
1035 related, but neither was related to 3), suggesting at least two separate inhibitory functions of
1036 which the Stroop task probes only one. Furthermore, as noted above, no task is ever a "pure"
1037 measure of a given function, but always includes additional processes. In the current study,
1038 the Stroop task was chosen as the means of assessing inhibition because it is widely used in
1039 the literature, allowing us to directly compare our findings to those of other studies, and
1040 because of the questions it has raised surrounding cross-modal comparability and potential
1041 non-inhibitory confounds, allowing us to explore the ability of alternative scoring methods to
1042 address these issues. However, a different choice of task is likely to have tapped different
1043 inhibitory functions and/or different additional processes, and therefore produced different
1044 relationships both across task modalities and also with SiN perception. Nevertheless, this
1045 only confirms our view that any given “inhibition” task does not necessarily provide a
1046 reliable measure of general inhibitory abilities, and that care must be taken when selecting
1047 both tasks and scoring systems.

1048

1049 One important limitation of this study is its restricted pool of participants – we only tested
1050 older adults with mild hearing loss. Nevertheless, within these confines, participant variables
1051 had a considerable range: 25 years in age and 30dB in hearing loss. This is important to keep
1052 in mind when examining data from other, more restrictive, samples, since the range defines
1053 the potential size of the modulating effect. How the relationships found in this study
1054 generalise to other groups of listeners needs to be investigated in further work. The number of
1055 participants used in the study was also relatively small, which may mean that individual
1056 variability and/or measurement error obscured effects. Replication with larger sample sizes is
1057 therefore desirable before firm conclusions are drawn.

1058

1059 It is also worth observing that the background masker used in the SiN task was speech-
1060 modulated noise, which contained no linguistic information. If the SiN stimuli had been
1061 presented in a speech masker, such as few-talker babble in which individual words were
1062 perceptible, then the observed relationships between SiN and Stroop scores might have been
1063 different. For example, it is possible that such SiN stimuli would demand a higher level of
1064 inhibition than those used here, since listeners would have to suppress not just noise but also
1065 lexical information, including the lexical neighbourhood of masker words (Helfer & Jesse,
1066 2015). However, it is hard to predict how this might have affected the Stroop/SiN relationship
1067 given the complex pattern of results obtained here. Finally, as discussed above, a further
1068 limitation of the study occurs in the form of the vSI_{res} measure, and in particular its reliance
1069 on a relationship based on the sample data rather than population norms. The predictive

1070 relationship between DI and C_i used to derive the vSI_{res} measure relies on the performance of
1071 the sample, which may not be representative of the wider population. If the vSI_{res} measure is
1072 considered to be useful, then future work should seek to establish an independent gold-
1073 standard relationship between C_i and DI.

1074

1075 **8 Abbreviations**

1076

1077 vSI_{raw} = traditional Stroop interference score (visual)

1078 aSI_{raw} = traditional Stroop interference score (auditory)

1079 C_i = overall colour naming time, incongruent condition (visual)

1080 C_n = overall colour naming time, neutral condition (visual)

1081 R_n = overall word reading time, neutral condition (visual)

1082 DI = dimensional imbalance (visual)

1083 y_{C_i} = observed C_i scores

1084 \hat{y}_{C_i} = predicted C_i scores

1085 vSI_{res} = residuals resulting from the difference between observed and predicted C_i scores
1086 (visual)

1087 aRT_i = average single-trial reaction time, incongruent condition (auditory)

1088 aRT_n = average single-trial reaction time, neutral condition (auditory)

1089 aRT_c = average single-trial reaction time, congruent condition (auditory)

1090 vSI_{norm} = normalised Stroop interference score (visual)

1091 aSI_{norm} = normalised Stroop interference score (auditory)

1092 $aSI_{norm\ delta}$ = normalised delta score (auditory)

1093

1094 **9 Funding**

1095

1096 This work was supported by the Medical Research Council [U135097128] and the
1097 Biotechnology and Biological Sciences Research Council (BB/K021508/1).

1098

1099 **10 Acknowledgements**

1100

1101 A subset of the reported data has appeared in published conference proceedings (Heinrich &
1102 Knight, 2016).

1103

1104 **11 Author contributions**

1105

1106 AH and SK designed the study and collected the data. SK analyzed the data. AH and SK
1107 interpreted the data. SK wrote, and AH contributed to, the manuscript and both contributed to
1108 the critical discussions. All authors approved the final version of the manuscript for
1109 publication.

1110 **12** **References**

- 1111
- 1112 Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting:
1113 retrieval dynamics in long-term memory. *Journal of Experimental Psychology:*
1114 *Learning, Memory, and Cognition*, 20(5), 1063-1087.
- 1115 Anstey, K. J., Dain, S., Andrews, S., & Drobny, J. (2002). Visual abilities in older adults
1116 explain age-differences in Stroop and fluid intelligence but not face recognition:
1117 Implications for the vision-cognition connection. *Aging, Neuropsychology, and*
1118 *Cognition*, 9(4), 253-265.
- 1119 Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual
1120 differences in cognitive ability? A survey of twenty experimental studies with normal
1121 and hearing-impaired adults. *International Journal of Audiology*, 47(sup2), S53-S71.
- 1122 Akutsu, H., Legge, G. E., Ross, J. A., & Schuebel, K. J. (1991). Psychophysics of reading—
1123 X. Effects of age-related changes in vision. *Journal of Gerontology*, 46(6), P325-
1124 P331.
- 1125 Aslan, A., & Bäuml, K. H. T. (2011). Individual differences in working memory capacity
1126 predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning,*
1127 *Memory, and Cognition*, 37(1), 264-269.
- 1128 Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review*
1129 *of psychology*, 63, 1-29.
- 1130 Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for color-word Stroop effects
1131 in aging: A meta-analysis. *Aging, Neuropsychology, and Cognition*, 16(5), 505-534.
- 1132 Ben-David, B. M., & Schneider, B. A. (2010). A sensory origin for color-word Stroop effects
1133 in aging: Simulating age-related changes in color-vision mimics age-related changes
1134 in Stroop. *Aging, Neuropsychology, and Cognition*, 17(6), 730-746.
- 1135 Bierer, J. A., Spindler, E., Bierer, S. M., & Wright, R. (2016). An examination of sources of
1136 variability across the consonant-nucleus-consonant test in Cochlear Implant listeners.
1137 *Trends in Hearing*, 20, <http://doi.org/10.1177/2331216516646556>
- 1138 Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization
1139 of a test of speech perception in noise. *Journal of Speech, Language, and Hearing*
1140 *Research*, 27(1), 32-48.
- 1141 Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-
1142 native listeners: Talker-, listener-, and item-related factors. *The Journal of the*
1143 *Acoustical Society of America*, 106(4), 2074-2085.
- 1144 Burke, D. M. (1997). Language, aging, and inhibitory deficits: Evaluation of a theory. *The*
1145 *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(6),
1146 P254-P264.
- 1147 Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life
1148 span. *Acta psychologica*, 86(2), 109-197.
- 1149 Committee on Hearing and Bioacoustics and Biomechanics (CHABA) (1988). Speech
1150 understanding and aging. *The Journal of the Acoustical Society of America*, 83, 859-
1151 895.

Stroop Interference and Speech Intelligibility

- 1152 Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent
1153 inhibition model. *Journal of experimental Psychology: general*, 123(4), 354-373.
- 1154 Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other
1155 psycholinguistic statistics. *Behavior research methods*, 37(1), 65-70.
- 1156 Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict
1157 audiovisual speech perception. *Psychology and aging*, 30(3), 634-646
- 1158 Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
- 1159 Dywan, J., & Murphy, W. E. (1996). Aging and inhibitory control in text comprehension.
1160 *Psychology and Aging*, 11(2), 199-206.
- 1161 Eisenberg, L. S., Martinez, A. S., Holowecky, S. R., & Pogorelsky, S. (2002). Recognition of
1162 lexically controlled words and sentences by children with normal hearing and children
1163 with cochlear implants. *Ear and Hearing*, 23(5), 450-462.
- 1164 Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference
1165 control functions: a latent-variable analysis. *Journal of experimental psychology:*
1166 *General*, 133(1), 101.
- 1167 Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of
1168 PRESTO: A new high-variability sentence recognition test. *Journal of the American*
1169 *Academy of Audiology*, 24(1), 26-36.
- 1170 Green, E. J., & Barber, P. J. (1981). An auditory Stroop effect with judgments of speaker
1171 gender. *Perception & Psychophysics*, 30(5), 459-466.
- 1172 Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review
1173 and a new view. *Psychology of learning and motivation*, 22, 193-225.
- 1174 Heinrich, A. & Knight, S. (2016). The contribution of auditory and cognitive factors to
1175 intelligibility of words and sentences in noise. In P. van Dijk et al. (Eds.), *Physiology,*
1176 *Psychoacoustics and Cognition in Normal and Impaired Hearing, Advances in*
1177 *Experimental Medicine and Biology* 894.
- 1178 Heinrich, A., Knight, S., Young, M., Moorhouse, R., & Barry, J. (2014) Assessing the effects
1179 of semantic context and background noise for speech perception with a new British
1180 English sentences set test (BESST). Paper presented at the British Society of
1181 Audiology Annual Conference, Keele University, UK
- 1182 Helfer, K. S., & Freyman, R. L. (2014). Stimulus and listener factors affecting age-related
1183 changes in competing speech perception. *The Journal of the Acoustical Society of*
1184 *America*, 136(2), 748-759.
- 1185 Helfer, K. S., & Jesse, A. (2015). Lexical influences on competing speech perception in
1186 younger, middle-aged, and older adults. *The Journal of the Acoustical Society of*
1187 *America*, 138(1), 363-376.
- 1188 Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing
1189 speech in older adults. *Aging, Neuropsychology, and Cognition*, 19(6), 741-758.
- 1190 Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of*
1191 *the Acoustical society of America*, 49(2B), 467-477.

Stroop Interference and Speech Intelligibility

- 1192 Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: a
1193 strong connection. *Psychology and aging*, 9(3), 339-355.
- 1194 Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation
1195 model. *Ear and hearing*, 19(1), 1-36.
- 1196 Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on
1197 audiovisual word recognition by adults with cochlear implants. *Journal of Speech,
1198 Language, and Hearing Research*, 46(2), 390-404.
- 1199 MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative
1200 review. *Psychological bulletin*, 109(2), 163-203.
- 1201 Macleod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual
1202 speech-reception thresholds for sentences in noise: Rationale, evaluation, and
1203 recommendations for use. *British journal of audiology*, 24(1), 29-43.
- 1204 Melara, R. D., & Algom, D. (2003). Driven by information: a tectonic theory of Stroop
1205 effects. *Psychological review*, 110(3), 422-471.
- 1206 Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function
1207 of the context of the test materials. *Journal of experimental psychology*, 41(5), 329-
1208 335.
- 1209 Morgan, A. L., & Brandt, J. F. (1989). An auditory Stroop effect for pitch, loudness, and
1210 time. *Brain and language*, 36(4), 592-603.
- 1211 Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by
1212 young children and older adults. *The Journal of the Acoustical Society of America*,
1213 87(6), 2705-2715.
- 1214 Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults
1215 listen to and remember speech in noise. *The Journal of the Acoustical Society of
1216 America*, 97(1), 593-608.
- 1217 Ridderinkhof, R. K. (2002). Micro-and macro-adjustments of task set: Activation and
1218 suppression in conflict tasks. *Psychological research*, 66(4), 312-323.
- 1219 Ridderinkhof, K. R., van den Wildenberg, W. P., Wijnen, J., & Burle, B. (2004). Response
1220 inhibition in conflict tasks is revealed in delta plots. *Cognitive neuroscience of
1221 attention*, 369-377.
- 1222 Roelofs, A., Piai, V., & Garrido Rodriguez, G. (2011). Attentional inhibition in bilingual
1223 naming performance: evidence from delta-plot analyses. *Frontiers in psychology*, Vol
1224 2, 184 <http://dx.doi.org/10.3389/fpsyg.2011.00184>
- 1225 Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The
1226 Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 50(6),
1227 P297-P306.
- 1228 Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across
1229 measures of inhibition: An investigation of the construct validity of inhibition in older
1230 adults. *Neuropsychologia*, 40(6), 605-619.
- 1231 Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for
1232 cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of*

Stroop Interference and Speech Intelligibility

- 1233 *aging and cognition (2nd edition)* (pp. 155-219). Mahwah, NJ: Lawrence Erlbaum
1234 Associates Publishers
- 1235 Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution
1236 to age-related declines in spoken-word recognition. *Psychology and aging*, *11*(2),
1237 333-341.
- 1238 Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word
1239 recognition in young and older adults: The interaction of lexical competition and
1240 semantic context. *Psychology and aging*, *14*(3), 458-472.
- 1241 Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve
1242 concept. *Journal of the International Neuropsychological Society*, *8*(3), 448-460.
- 1243 Stoltzfus, E. R., Hasher, L., & Zacks, R. T. (1996). Working memory and aging: Current
1244 status of the inhibitory view. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H.
1245 Logie, E. R. Stoltzfus & R. T. Zacks (Eds), *Working memory and human cognition*
1246 (pp. 66-88). New York, NY: Oxford University Press.
- 1247 Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental*
1248 *psychology*, *18*(6), 643-662.
- 1249 Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech, Language,*
1250 *and Hearing Research*, *28*(3), 455-462.
- 1251 Surprenant, A. M. & Neath, I. (2009). The Impurity Principle. In *Principles of Memory* (pp.
1252 85-101). New York, NY: Psychology Press.
- 1253 Taler, V., Aaron, G.P., Steinmetz, L.G., & Pisoni, D.B. (2010). Lexical neighborhood density
1254 effects on spoken word recognition and production in healthy aging. *Journal of*
1255 *Gerontology: Psychological Sciences*, *65B*(5), 551–560.
- 1256 Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of
1257 meta-analyses. *Neuroscience & Biobehavioral Reviews*, *26*(7), 849-857.
- 1258 von Hippel, W. (2007). Aging, executive functioning, and social control. *Current directions*
1259 *in psychological science*, *16*(5), 240-244.
- 1260 Werner, J. S., & Steele, V. G. (1988). Sensitivity of human foveal color mechanisms
1261 throughout the life span. *JOSA A*, *5*(12), 2122-2130.
- 1262 Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on
1263 comprehension of spoken language. *Journal of the American Academy of Audiology*,
1264 *18*(7), 548-558.

1265 **Tables and figure captions**

1266 Figure 1: Mean PTA thresholds as a function of frequency. Bars indicate standard deviation.

1267 Figure 2: Delta plot showing each individual's aSI_{delta} scores across the five quintiles

1268 Table 1: Lexical information for word stimuli

		LOW WF LOW ND	LOW WF HIGH ND	HIGH WF LOW ND	HIGH WF HIGH ND
WF	Max	9879	8958	41358	62803

Stroop Interference and Speech Intelligibility

	Min	106	117	10152	10029
ND	Max	18	38	18	35
	Min	2	19	2	19

1269

1270 Table 2: Mean scores in the 6 different SiN conditions

Sentences	Semantic predictability		Easy SNR (-4dB)	Hard SNR (-7dB)
	HP		0.88	0.73
	LP		0.57	0.41
Words	Word Frequency	Neighborhood Density	Easy SNR (+1dB)	Hard SNR (-2dB)
	High WF	High ND	0.71	0.58
	High WF	Low ND	0.82	0.76
	Low WF	High ND	0.72	0.64
	Low WF	Low ND	0.67	0.60

1271

1272 Table 3: Summary of LMMs assessing relationship of visual Stroop scores to sentence
1273 perception

Outcome variable: sentences

Scoring method: vSI_{raw}			
Stimulus-based predictors: semantic predictability (high/low), SNR (high/low)			
AIC value	ME	Interaction(s) involving Stroop	Description
Listener-based predictors: Stroop			
1426.747	N	N	N/A
Listener-based predictors: Stroop, PTA			
1394.693	N	N	N/A
Scoring method: vSI_{res}			
Stimulus-based predictors: semantic predictability (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
1429.328	N	(1) $vSI_{res} * Pred * SNR$	(1) At the high (easy) SNR, the slope predicting SiN performance from Stroop interference is positive for HP sentences and negative for LP. At the low (hard SNR), the slope is negative for HP and positive for LP.
Listener-based predictors: Stroop, PTA			
1396.551	N	(1) $vSI_{res} * Pred * SNR$	(1) As above.

1274

1275 Table 4: Summary of LMMs assessing relationship of visual Stroop scores to word
1276 perception

Outcome variable: words

Scoring method: vSI_{raw}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
AIC value	ME	Interaction(s) involving Stroop	Description

Stroop Interference and Speech Intelligibility

Listener-based predictors: Stroop			
2708.973	N	(1) $vSI_{raw} * ND$	(1) The slope predicting SiN performance from Stroop interference is negative overall, and most strongly for words with low neighborhood density (ND).
Listener-based predictors: Stroop, PTA			
2695.725	N	(1) $vSI_{raw} * ND$ (2) $vSI_{raw} * SNR * ND * PTA$	(1) The slope predicting SiN performance from Stroop interference is negative overall, and most strongly for words with low neighborhood density (ND). (2) For those with poor PTA, the slope predicting SiN performance from Stroop interference is negative and stronger for low ND words. For those with good PTA, the slope is positive for high ND words and negative for low ND words at the easier SNR, and approaches zero for both ND categories at the harder SNR.
Scoring method: vSI_{res}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
2712.168	N	N	N/A
Listener-based predictors: Stroop, PTA			
2691.369	N	(1) $vSI_{res} * ND$ (2) $vSI_{res} * ND * PTA$	(1) The slope predicting SiN performance from Stroop interference is negative overall, and most strongly for words with low neighborhood density (ND). (2) For those with good PTA, the slope predicting SiN performance from Stroop interference is negative for both ND categories. For those with poor PTA, the slope is more strongly negative for low ND words and approaches zero for high ND words.

1277

1278 Table 5: Summary of LMMs assessing relationship of visual Stroop scores to all SiN
1279 perception (combined dataset)

Outcome variable: speech (combined dataset)

Scoring method: vSI_{raw}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
1266.480	N	(1) $vSI_{raw} * Type$	(1) The slope predicting SiN performance from Stroop interference is negative for words and mildly positive for sentences.
Listener-based predictors: Stroop, PTA			
1236.257	N	(1) $vSI_{raw} * Type$	(1) As above.

Stroop Interference and Speech Intelligibility

Scoring method: vSI_{res}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
Listener-based predictors: Stroop			
1270.403	N	N	N/A
Listener-based predictors: Stroop, PTA			
1239.501	N	N	N/A

1280

1281 Table 6: Summary of LMMs assessing relationship of auditory Stroop scores to sentence
1282 perception

Outcome variable: sentences

Scoring method: aSI_{raw}			
Stimulus-based predictors: predictability (high/low), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
1459.850	N	N	N/A
Listener-based predictors: Stroop, PTA			
1428.302	N	N	N/A
Scoring method: aSI_{norm}			
Stimulus-based predictors: predictability (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
1456.132	Y	N	N/A
Listener-based predictors: Stroop, PTA			
1427.957	N	(1) $aSI_{norm} * Pred * SNR * PTA$	(1) For those with good PTA, the slope predicting SiN performance from Stroop interference is positive for HP sentences at the easier SNR and LP sentences at the harder SNR, and approaches zero elsewhere. For those with poor PTA, the slope is positive for HP sentences at the harder SNR and LP sentences for the easier SNR, and approaches zero elsewhere.

1283

1284 Table 7: Summary of LMMs assessing relationship of auditory Stroop scores to word
1285 perception

Outcome variable: words

Scoring method: aSI_{raw}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
2776.946	N	(1) $aSI_{raw} * SNR$	(1) The slope predicting SiN performance from Stroop interference is positive overall, and more strongly so at the harder SNR.
Listener-based predictors: Stroop, PTA			

Stroop Interference and Speech Intelligibility

2759.515	N	(1) $aSI_{raw} * SNR$	(1) As above.
Scoring method: aSI_{norm}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
2771.321	Y	(1) $aSI_{norm} * SNR$	(1) The slope predicting SiN performance from Stroop interference is positive in both conditions, and more strongly so at the harder SNR.
Listener-based predictors: Stroop, PTA			
2755.034	N	(1) $aSI_{norm} * SNR$	(1) As above.

1286

1287 Table 8: Summary of LMMs assessing relationship of auditory Stroop scores to all SiN
1288 perception (combined dataset)

Outcome variable: speech (combined dataset)

Scoring method: aSI_{raw}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
1289.565	N	(1) $aSI_{raw} * SNR$	(1) The slope predicting SiN performance from Stroop interference is positive overall, and more strongly so for the harder SNR.
Listener-based predictors: Stroop, PTA			
1260.049	N	(1) $aSI_{raw} * SNR$	(1) As above.
Scoring method: aSI_{norm}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
Listener-based predictors: Stroop			
1285.224	Y	(1) $aSI_{norm} * SNR$	(1) The slope predicting SiN performance from Stroop interference is positive overall, and more strongly so for the harder SNR.
Listener-based predictors: Stroop, PTA			
1256.700	Y	(1) $aSI_{norm} * SNR$	(1) As above.

1289

1290 Table 9: Summary of LMMs assessing relationship of auditory Stroop delta scores to
1291 sentence perception

Outcome variable: sentences

Scoring method: $aSI_{ndeltaQ5}$			
Stimulus-based predictors: predictability (high/low), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
1493.843	N	(1) $aSI_{ndeltaQ5} * Pred * SNR$	(1) The slope predicting SiN perception from Stroop interference is positive for LP sentences at the easier SNR and HP

Stroop Interference and Speech Intelligibility

			sentences at the harder SNR, and approaches zero elsewhere.
Listener-based predictors: Stroop, PTA			
1457.746	N	(1) $aSI_{\text{ndeltaQ5}} * \text{Pred} * \text{SNR}$	(1) As above.
Scoring method: aSI_{ndeltaQ1}			
Stimulus-based predictors: predictability (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
1491.747	N	N	N/A
Listener-based predictors: Stroop, PTA			
1458.472	N	N	N/A

1292

1293 Table 10: Summary of LMMs assessing relationship of auditory Stroop delta scores to word
1294 perception

Outcome variable: words

Scoring method: aSI_{ndeltaQ5}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
2827.234	Y	(1) $aSI_{\text{ndeltaQ5}} * \text{SNR}$	(1) The slope predicting SiN perception from Stroop interference is positive overall, and more strongly so at the harder SNR.
Listener-based predictors: Stroop, PTA			
2807.669	Y	(1) $aSI_{\text{ndeltaQ5}} * \text{SNR}$	(1) As above.
Scoring method: aSI_{ndeltaQ1}			
Stimulus-based predictors: word frequency (high/low), neighborhood density (high/low), SNR (high/low)			
Listener-based predictors: Stroop			
2833.745	N	N	N/A
Listener-based predictors: Stroop, PTA			
2817.638	N	N	N/A

1295

1296 Table 11: Summary of LMMs assessing relationship of auditory Stroop delta scores to all SiN
1297 perception (combined dataset)

Outcome variable: speech (combined dataset)

Scoring method: aSI_{ndeltaQ5}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
<i>AIC value</i>	<i>ME</i>	<i>Interaction(s) involving Stroop</i>	<i>Description</i>
Listener-based predictors: Stroop			
1321.151	N	(1) $aSI_{\text{ndeltaQ5}} * \text{Type}$ (2) $aSI_{\text{ndeltaQ5}} * \text{SNR}$	(1) The slope predicting SiN perception from Stroop interference is positive overall, and more strongly so for words. (2) The slope predicting SiN perception from Stroop interference is

Stroop Interference and Speech Intelligibility

			positive overall, and more strongly so for the harder SNR.
Listener-based predictors: Stroop, PTA			
1282.466	N	(1) $aSI_{\text{ndeltaQ5}} * SNR$ (2) $aSI_{\text{ndeltaQ5}} * \text{Type} * PTA$	(1) As above. (2) The positive slope predicting SiN performance from Stroop interference is stronger for sentences for those with good PTA and stronger for words for those with poor PTA.
Scoring method: aSI_{ndeltaQ1}			
Stimulus-based predictors: type (sentences/words), SNR (high/low)			
Listener-based predictors: Stroop			
1325.809	N	N	N/A
Listener-based predictors: Stroop, PTA			
1294.172	N	(1) $aSI_{\text{ndeltaQ1}} * \text{Type} * PTA$	(1) For those with good PTA, the slope predicting SiN perception from Stroop interference is negative and stronger for words. For those with poor PTA, the slope is negative and stronger for sentences.

1298

1299 Table 12: Intercorrelations of all Stroop scoring systems (visual and auditory)

1300

	vSI_{raw}	vSI_{res}	aSI_{raw}	aSI_{norm}	aSI_{ndeltaQ5}	aSI_{ndeltaQ1}
vSI_{raw}	-					
vSI_{res}	.763**	-				
aSI_{raw}	-.013	.050	-			
aSI_{norm}	-.009	.008	.953**	-		
aSI_{ndeltaQ5}	-.265	-.213	.815**	.850**	-	
aSI_{ndeltaQ1}	.208	.117	.384**	.406**	.202	-

1301

Figure 01.JPEG

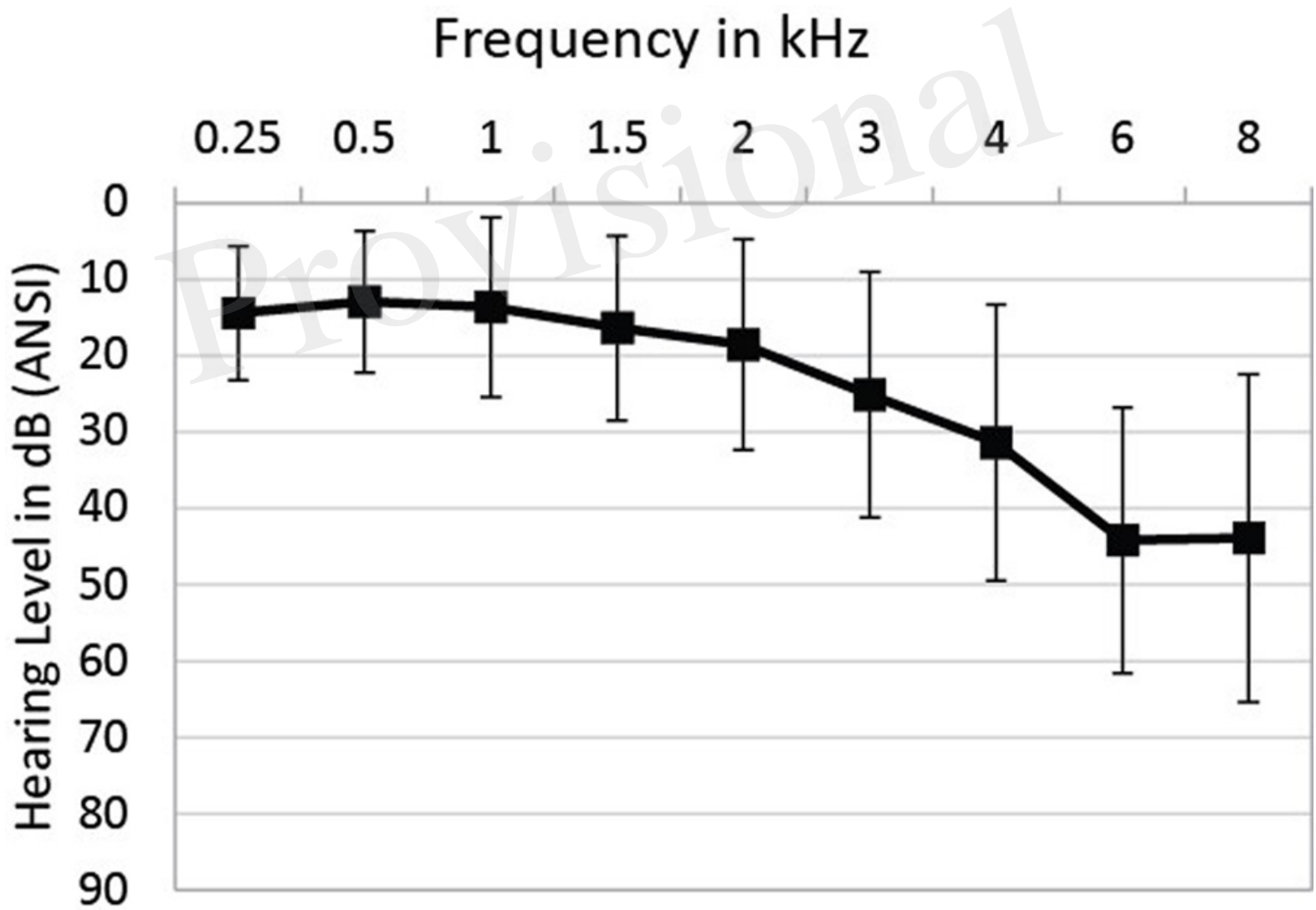


Figure 02.JPEG

