



HAL
open science

Irrelevant speech effect in open plan offices: Comparison of two models explaining the decrease in performance by speech intelligibility and attempt to reduce interindividual differences of the mental workload by task customisation

Krist Kostallari, Etienne Parizet, Patrick Chevret, Jean-Noël Amato, Edith Galy

► To cite this version:

Krist Kostallari, Etienne Parizet, Patrick Chevret, Jean-Noël Amato, Edith Galy. Irrelevant speech effect in open plan offices: Comparison of two models explaining the decrease in performance by speech intelligibility and attempt to reduce interindividual differences of the mental workload by task customisation. *Applied Acoustics*, 2020, 161, pp.107180. 10.1016/j.apacoust.2019.107180. hal-03212357

HAL Id: hal-03212357

<https://hal.science/hal-03212357v1>

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Irrelevant speech effect in open plan offices: comparison of two models explaining the
2 decrease in performance by speech intelligibility and attempt to reduce interindividual
3 differences of the mental workload by task customisation

4

5 Kostallari K.*^{a,b}, Parizet E.^b, Chevret P.^a, Amato J.-N.^a, Galy E.^c

6 ^a INRS – The French National Research and Safety Institute for the Prevention of
7 Occupational Accidents and Diseases, Vandœuvre-lès-Nancy, France

8 ^b Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, F-69621 Villeurbanne, France

9 ^c LAPCOS – Laboratory of Anthropology, Clinical, Cognitive and Social Psychology, Nice,
10 France

11

12 **Abstract**

13
14 For employees working in open plan offices, speech noise is now recognised as the main
15 source of disturbance. In the literature, a series of studies on open spaces have found that
16 the increased speech intelligibility happens to decrease the performance during a short-
17 term memory task. The theoretical model proposed by Hongisto relates the decreased
18 performance (DP) to the Speech Transmission Index (STI). The model predicts that for STI
19 values from 0.7 to 1, which correspond to a speech signal with an intelligibility of almost
20 100%, the DP remains constant at 7%, but few studies have examined the high end of this
21 range. Here, we investigated the relationship between the DP and the STI by varying the STI
22 up to 0.9. Fifty-five subjects between 25- and 59-years-of-age participated in the
23 experiment. Subjects performed a short-term memory task in silence and in four different
24 sound conditions (STI from 0.25 to 0.9). The task itself was personalised by prior
25 measurement of mnemonic span. It was thus possible to define two different cognitive
26 loads (low/high) based on the mnemonic span value determined for each subject. Subjects
27 subjectively assessed the mental workload and sound annoyance at the end of each short-
28 term memory task in each sound condition. Significant interactions between STI and DP,
29 mental workload and sound annoyance were found. In addition, the age of subjects
30 correlated significantly with their performance during the short-term memory task.

31
32 **Keywords:** Open plan offices, Speech Intelligibility, Performance, Mental workload, Sound
33 annoyance

34

35

36 **1. Introduction**

37 An open plan office can be defined as a workspace designed to host more than five people
38 without complete separation between the workstations [1]. These spaces are economical
39 for the employer, who also **presume**, even if there is no scientific proof, that they facilitate
40 teamwork. Nevertheless, in open plan offices noise is considered to be a major source of
41 disturbance by employees [2-4]. By analysing types of noise, field studies indicated that
42 speech generated by other occupants of the space was the most disturbing type of noise [5-
43 7].

44 The perturbing effect of speech is linked to Irrelevant Speech Effect (ISE), which describes
45 how a vocal signal can perturb a task involving short-term memory [8, 9]. In laboratory-
46 based experiments, the ISE was particularly observed when participants were asked to
47 retain lists of words or digits presented visually, called a serial recall task, while
48 simultaneously exposed to speech. If the participants were exposed to speech, their
49 performance degraded significantly during the serial recall task [10, 11]. Several
50 experiments have attempted to identify the parameters of speech causing this effect (see
51 [12] for review). It appears that the effect of the nonstationary nature (in amplitude and
52 frequency) of the signal was found to be significant. In addition, a non-speech signal that
53 presents similar characteristics to speech also disrupts the accomplishment of the task.
54 Researchers therefore considered that the ISE should be redefined as an Irrelevant Sound
55 Effect.

56 However, in an open-plan office, nonstationary signals that may be overheard by an
57 occupant will mainly be speech. This fact justified the search for intelligibility indicators,
58 which could predict the decrease in performance created by these signals. The Speech

59 Transmission Index (STI) developed by Steeneken and Houtgast (1980) [13] is commonly
60 used to assess this intelligibility. This indicator can be used to quantify the degradation of a
61 speech signal as it travels from the source to the listener’s ear. The STI takes into account
62 the acoustical properties of the room, the hearing threshold of the listener and spectral
63 masking phenomena. It ranges between 0 and 1. Intelligibility is very poor at STI values
64 between 0 and 0.2 or 0.3 and very good at values greater than 0.7 or 0.8 (rated by IEC
65 60268-16 standard [14] as “Not difficult: no effort is required, equivalent to a relaxed listening
66 condition).

67 Several studies investigated the relation between intelligibility of a speech signal in
68 different background noises, characterised using the STI, and performance during a short-
69 term memory task. By observing the results of several studies, Hongisto (2005) [15]
70 proposed a model linking the STI and the decrease in performance (DP), defined as the
71 difference between the percentage of errors made in the sound condition and that
72 measured in a control condition, generally silence:

$$DP = Perf_{control} - Perf_{noise} \quad (1)$$

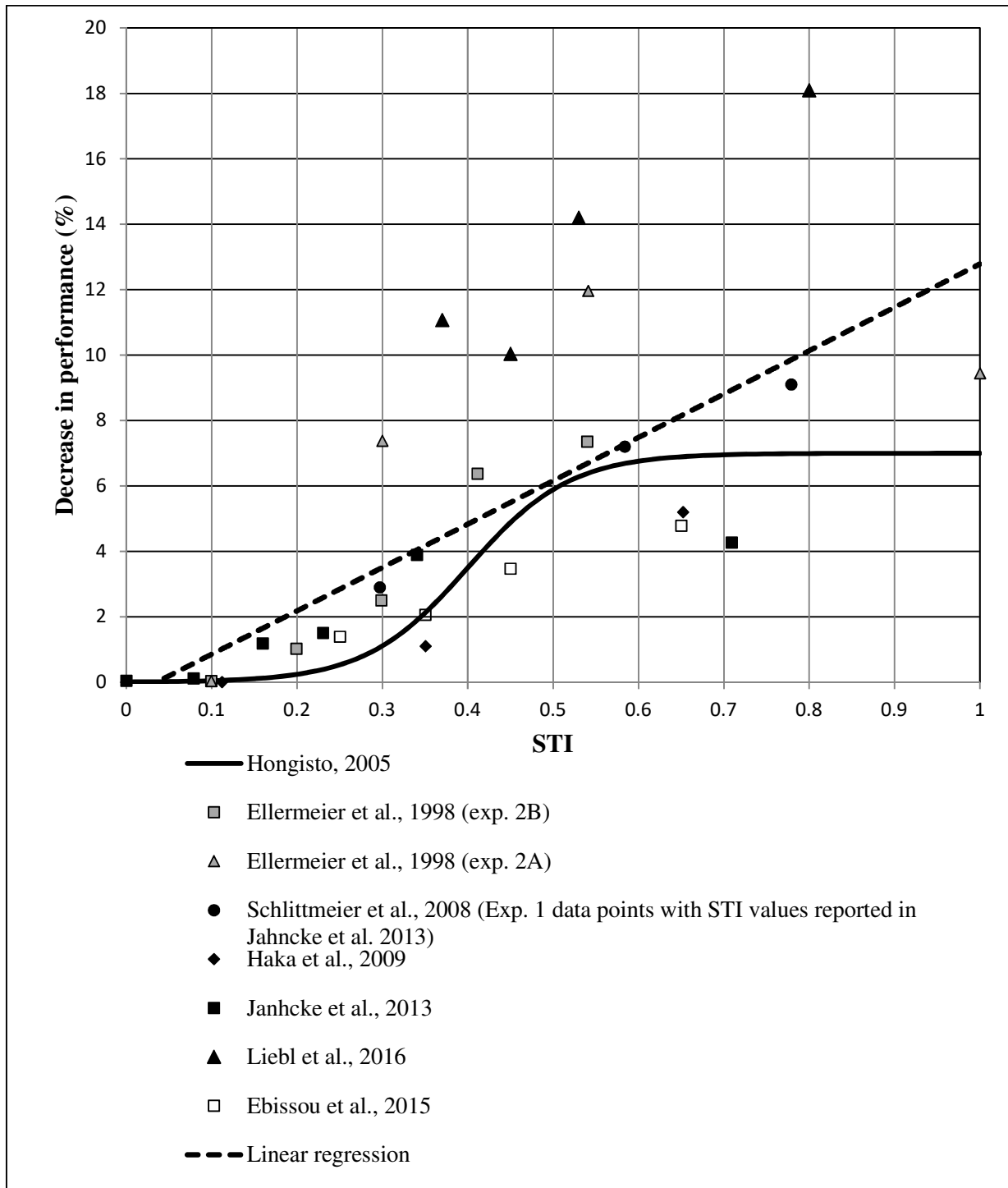
73 The following model was proposed:

$$DP = \frac{-7}{1 + \exp\left(\frac{STI - 0.4}{0.06}\right)} + 7 [\%] \quad (2)$$

74 Based on this relation, illustrated in figure 1 by the continuous line, for STI values between
75 0.25 and 0.6, the DP will be very sensitive to STI. Beyond this value, the DP will remain
76 almost constant at 7%. According to Hongisto, this ceiling value of 7% may differ depending
77 on the type of task used in the experiment.

78 This relation between STI and DP was taken in consideration in the ISO 3382-3 standard
79 (2012) [16] to define the indicators used to assess noise quality in open-plan offices. Among
80 the assessment indicators, two distances were defined and were directly linked to the STI:
81 the distraction distance and the privacy distance. The distraction distance represents the
82 distance between the speaker and the listener from which the STI remains less than 0.5. For
83 a distance less or equal to the distraction distance, the speaker's voice is considered to
84 remain intelligible, but its effect on the work performance of the other occupants will
85 remain acceptable. According to Virjonen et al. (2009) [17], the value $STI = 0.5$ is selected to
86 limit effects of speech on work performance. The other distance, the privacy distance, can
87 be defined as the distance from which an STI value of 0.2 is reached (speech is no longer
88 intelligible). According to Virjonen et al. (2009) [17], the value $STI = 0.2$ is selected to avoid
89 any effect of speech on work performance. However, there is no evidence for defining target
90 values for the STI and the two distances. These two values were chosen without considering
91 the variability of the experimental data represented in figure 1 [15, 18-23]. These data are
92 taken from the articles discussing the relation between decrease in performance and STI.
93 From these data, no major variations of the DP are visible between $STI = 0.2$ and $STI = 0$.
94 Thus, the intelligibility effect on the DP can be considered negligible in this range. In
95 contrast, for $STI = 0.5$, significant variability between studies is observed (the DP varies
96 between 2% and 14%). As a result, this value cannot be defined as a "target value" used for
97 the definition of the distraction distance. In addition, studies using STI values lower than 0.7
98 may have used different protocols which can explain this variability between the
99 experiments of the figure 1. For example, Jahncke et al. (2013) [21] asked participants to
100 remember words, whereas in other experiments, lists of digits were used. Schlittmeier et al.

101 (2008) [23] and Liebl et al. (2016) [22] applied a retention phase of 10 seconds, whereas
102 participants could give their responses as soon as the last item had been presented in the
103 protocols used by Ellermeier and Hellbrück (1998) [19] or Jahncke et al. (2013) [21]. This
104 can complicate the comparison of the results from these different experiments.
105



106
107

Figure 1. Relation between decrease in performance (DP) and STI for different experiments involving short-term memory tasks, the Hongisto model (continuous

108

109 line) and the fitted linear regression model (dashed line; $r^2=0.48$) for all data points
110 included in the experiments.

111 Moreover, some of the results presented in figure 1, such as those presented in Jahncke et
112 al. (2013) [21] fit to the plateau effect described by the Hongisto model [15]. However, for
113 others, no saturation was observed at high STI values. Thus, Schlittmeier et al. (2008) [23]
114 varied the STI between 0.3 and 0.8 (STI values calculated and reported in figure 1 of
115 Jahncke et al. (2013) [21]) and observed a constant increase of the DP. Later, Liebl et al.
116 (2016) [22] varied the STI between 0.37 and 0.8. They found a significant difference
117 between the given DP value for the condition where $STI = 0.8$ and all the other DP values.
118 These results show that Hongisto's model is not clearly validated using the results shown in
119 the literature.

120 Furthermore, in addition to the Hongisto's model [15], in figure 1 we have also represented
121 (dashed line) a linear regression linking the different experimental points. This linear
122 model was inspired by a model proposed by Schlittmeier et al. (2012) [24]. These authors
123 proposed that the decrease in performance is directly proportional to the fluctuation
124 strength defined by Zwicker and Fastl (2013) [25]. This indicator is strongly linked to
125 amplitude modulations of a signal. Its maximum is reached at a modulation frequency close
126 to 4 Hz, which corresponds to the mean syllabic flow of a constraint-free discussion. This
127 finding appears to confirm that the fluctuation strength increases with the intelligibility of
128 the speech signal, and thus with the STI. In the article [24] a linear regression with a
129 coefficient of determination of $r^2=0.55$, justifies the relation between the decrease in
130 performance and the fluctuation strength. In addition, while isolating only the speech sound

131 conditions the coefficient of determination of the linear regression, is still fairly low
132 ($r^2=0.51$).

133 Therefore, a first goal of this study was to provide further experimental values of the
134 decrease in performance in situations where speech intelligibility was found to have a
135 significant impact such as the indications given in the ISO 3382-3 standard (2012) [16] (for
136 STI values between 0.2 and 0.5), and to examine situations where speech intelligibility is
137 particularly high (STI greater than 0.7) so as to test the two models (linear or sigmoidal).

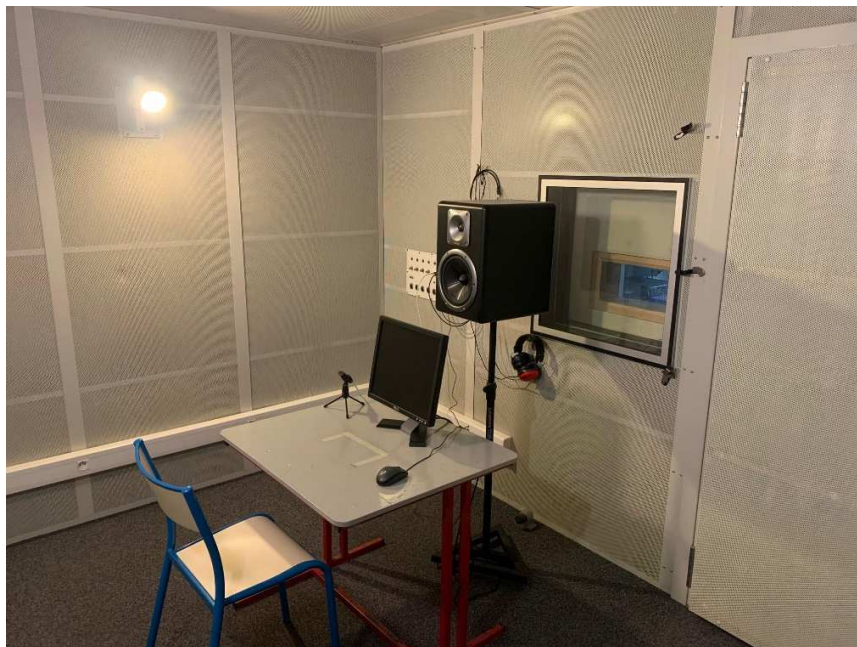
138 Additionally, beside performance, one can also be interested in the mental workload felt by
139 participants. According to Galy et al. (2015) [26], the mental workload helps to understand
140 how one person mobilizes his mental resources to perform a certain task. The mental
141 workload is a function of individual characteristics (cognitive capacities for example) and
142 environmental characteristics. In real open-plan offices, noise is likely to increase this
143 workload while people **try to maintain** their performance level. Generally speaking, it is
144 well known that people have different short-term memory capacities. Consequently, a
145 short-term memory task with a fixed number of elements to retain can be harder for some
146 participants than for others, because it generates a higher mental workload. Although this
147 variability has no influence on decrease of performance (see [27] for review), it can be
148 hypothesized that this it will increase the dispersion of individual workload measurement.
149 As an example, in Ebissou et al. (2015) [18] two groups could be established within the
150 panel of 57 participants, according to individuals' overall performances. For the high-
151 performers group, the mental workload measured using the NASA-RTLX (NASA Raw Task
152 Load Index) [28, 29], was also lower than the mental workload evaluated by the low-
153 performers group and this workload did not vary with STI. At the same time, workload

154 appeared to increase with STI for the low-performers group. Similarly, Brocolini et al.
155 (2016) [30] also found that the task was more demanding for the low-performers group.
156 Based on those results, a second goal of this study was to adapt the difficulty of the task to
157 each subject's short-term memory capacities by using a mnemonic span test, in order to
158 check whether this can make more apparent the influence of speech intelligibility to mental
159 workload.

160

161 2. Experimental procedures

162 The experiment performed for this study aimed to assess the sound-related decrease in
163 performance using a repeated short-term memory task in different sound conditions. The
164 experiment was conducted in the audiometric test booth at the Laboratory of Vibrations
165 and Acoustics (LVA) at INSA-Lyon. **The dimensions of the booth were 2.4 x 3.4 x 2.2 meters**
166 **and the global reverberation time was inferior to 100ms, starting from 250 Hz.**



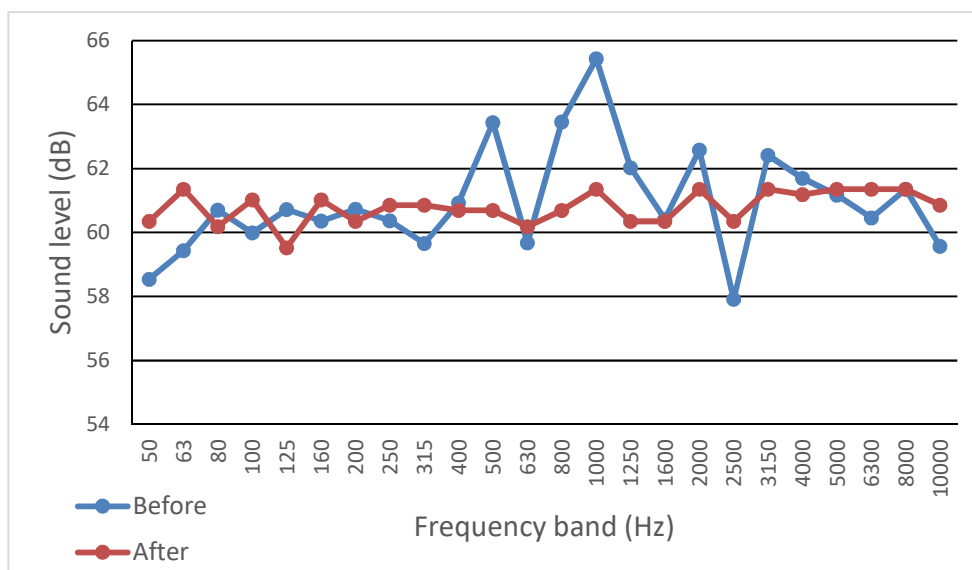
167

168 Figure 2. The audiometric test booth during the experiment

169

170 Inside the booth (figure 2), the participant was seated on a chair in front of a table. On the
171 table, a screen and mouse were placed and linked to the computer piloting the experiment.
172 The computer was located outside the room. The distance between the screen and the
173 subject's head was around 0.45 m. A loudspeaker (Tapco S8) was placed in front of the
174 subject/behind the screen. The loudspeaker was placed higher than the screen (around 1
175 m) and at a distance of around 0.7 m from the subject. A microphone, used to record
176 answers that were given orally by the subject, was placed on the table beside the screen.
177 **The loudspeaker in the room** was equalised in 1/3- octave bands to obtain a flat frequency
178 response at the subject's position. This equalisation was performed using a digital equaliser
179 (Behringer ULTRA-CURVE PRO DSP8024). Figure 3 shows the 1/3-octave bands spectrum
180 for a pink noise measured by an acoustic dummy-head (with the microphones position
181 inside the blocked ear canal) at the subject's position (mean sound levels of both ears of the
182 dummy-head), before and after equalisation.

183



184

185 Figure 3. Spectrum for a pink noise in the listener's position before (blue) and after
186 (red) equalization.

187

188 **2.1. Participants**

189 Fifty-five people, 26 women and 29 men, participated in this experiment. Their age ranged
190 between 25- and 59-years-of-age (mean: 41 years, median: 43 years). All participants were
191 informed of the nature and duration of this experiment. They were paid as compensation
192 for their participation in this study.

193 Prior to the test itself, participants' hearing was tested using an **audiometric** test to identify
194 any hearing loss. According to the BIAP (International Bureau for Audiophonology), hearing
195 loss is calculated as the mean of losses for frequencies from 500 Hz to 4000 Hz (see
196 <https://www.biap.org/en/>). By considering the best ear for each subject, 45 subjects were
197 considered as normal-hearing subjects, having hearing losses of less than or equal to 20 dB.
198 Six people had a hearing loss between 22 and 25 dB, and four had a hearing loss between
199 30 and 35 dB. According to the BIAP guidelines, these individuals present mild hearing
200 impairment. For these subjects, "Speech is perceived if the voice is normal, difficulties arise
201 if the voice is low or distant from the subject. Most of the daily life noises are perceived".
202 **After the experiment, we could not find any evidence that the different performers**
203 **depended on the hearing threshold of the participants.** Based on these results, for this
204 experiment, no subject was considered to have severe hearing loss. The data for all 55
205 subjects were therefore retained.

206

207 **2.2. Sound stimuli**

208 The participants performed the experiment in five different sound conditions to study the
209 effect of intelligibility during a short-term memory task. In four experimental conditions,
210 the task was performed in the presence of speech sounds including a more or less
211 intelligible speech signal, and in one condition the task was performed in silence. Four
212 different narrators were used for the intelligible conditions. The sound signals were
213 presented to the subject at an acoustic level of 55 dB(A), which is supposed to be
214 representative of the ambient noise level in open plan offices [6].

215 The signals for this experiment were composed of a stationary babble noise and perfectly
216 intelligible speech signals. The babble noise was created by superimposing 10 voices of the
217 same average level from the Hearing In Noise Test (HINT), and temporally reversing them.
218 The speech signals were taken from a free database of audiobooks on the Internet
219 (<http://www.litteratureaudio.com/>). Two female and two male narrators were selected based
220 on the quality of the recording. Then, the signals were cut and cleaned of any possible
221 noises or recording defects. Long pauses in the narration were also eliminated.

222 All the files were converted into *.wav* format. Finally, each signal lasted 9 minutes.

223 The speech signal for each narrator and the background noise were equalised by octave
224 band to obtain the omnidirectional version of Long-Term Average Speech Spectrum
225 (LTASS) as described in ISO 3382-3 standard (2012) [16]. This equalisation makes it
226 possible to apply an identical level of acoustic pressure for each octave band without
227 distinguishing between female and male narrators. Each extract was then added to the
228 background noise, maintaining the necessary signal/noise ratios at each octave band and
229 without any early decay time to produce the target STI values of 0.25; 0.45, values chosen in

230 the transition zone ($0.2 < STI < 0.5$) of Hongisto's model, and 0.75; 0.9, values chosen in the in
231 the zone of the hypothetical plateau of Hongisto's model ($0.7 < STI$).

232 The 16 signals created (four narrators \times four noise conditions corresponding to the STI
233 values for the experiment) were then normalised to obtain the same effective acoustic
234 pressure value.

235 Finally, using the data-acquisition card and the acoustic measurement system, the
236 amplitude of the signals was adjusted to obtain an overall noise level of 55 dB(A) at the
237 subject's position.

238

239 **2.3. Protocol**

240 The experiment took place in two phases. The first phase was performed in silence and
241 consisted in measuring the participants' mnemonic capacities and other psychological data
242 which could lead to variability between individual results. The second phase included the
243 short-term memory task, performed in one of the five sound conditions described above.
244 The total duration of the experiment was around one hour and 30 minutes (40 minutes for
245 the preliminary phase and 50 minutes for the main phase). One third of the participants
246 started the experiment at 10 AM, one third at 2 PM and one third at 4 PM, approximately.

247

248 **2.3.1. Preliminary phase**

249 This phase included the four following steps:

250 - *Measurement of mnemonic capacities* [31], i.e., the subject's maximal capacity to
251 memorise a certain number of elements and repeat them immediately after hearing them.

252 The participant had to retain lists of numbers which were presented orally without

253 background noise, at a pace of around one number per second. Immediately after the list
254 had been presented, the subject repeated the numbers that they had retained in reverse
255 order. The number of elements on the list increased every two lists (from two), until the
256 participant could no longer retain all of the numbers presented. The mnemonic value
257 therefore corresponds to the number of elements on the last list repeated correctly. This
258 value was used to customize the number of elements that the participants had to retain
259 during the short-term memory task, which is part of the second goal of this study.

260 - *STROOP task* [32], to assess the subject's sensitivity to interference effect due to the
261 presence of irrelevant information while performing a cognitive task. Sensitivity to this
262 interference effect was measured by performing the STROOP test in three stages:

- 263 • Firstly, the name of a colour (red, blue, green) was presented on the screen (font: 64-
264 point Helvetica) alongside three buttons corresponding to the three possibilities.
265 The subject had to click as fast as possible on the button corresponding to the word
266 displayed. This action was repeated 100 times successively.
- 267 • Secondly, coloured squares measuring 4 x 4 cm (red, green, blue) were displayed on
268 the screen, sequentially, beside the three boxes corresponding to the three
269 possibilities. The subject had to click as fast as possible on the button corresponding
270 to the colour of the square. This action was repeated 100 times successively.
- 271 • Thirdly, the name of a colour (red, blue, green) written in a different-coloured font
272 (red, blue, green) and three buttons corresponding to the three possibilities were
273 displayed on the screen. The font and button sizes were identical to those used in the
274 previous stages. The subject had to click as fast as possible on the button

275 corresponding to the colour in which the word was written. This action was
276 repeated 100 times successively.

277 The time required to complete each stage was measured. Finally, a weak interference
278 index was calculated as the ratio between the time to complete the first stage and the
279 time to complete the second stage, and a strong interference index was calculated based
280 on the ratio between the time to complete the third stage and the time to complete the
281 second stage. It was assumed that this sensitivity to this interference effect could play a
282 role on the cognitive load exerted to the participant by the short-term memory task.

283 - *The Thayer questionnaire* [33] was used to assess the participant's energetic arousal.
284 It consists of 20 questions. The participant was asked to answer these questions
285 initially before starting the short-term memory task and at the end of each sound
286 condition (main phase). The measurement of the energetic arousal during this study
287 is an attempt to better explain the high interindividual variability of the performance
288 during the short-term memory task.

289

290 **2.3.2. Main phase**

291 This phase of the experiment was designed to measure the effects of speech intelligibility
292 on performance and the annoyance induced during a short-term memory task.

293 Participants were asked to retain a series of words. Before each series, a cross, known as a
294 fixation cross, appeared in the centre of the screen for three seconds. This cross was
295 followed by a list of words which appeared one by one. The words were written in black on
296 white background, in 64-point Helvetica. Each word remained on the screen for one second.
297 The words followed each other without interruption. When the list was finished, the screen

298 remained white for three seconds (retention phase). When the screen displayed “Over to
299 you”, the subject had to repeat the words they remembered out loud, in no particular order.
300 The message “C’est à vous” (“Over to you”) remained displayed for 15 seconds, which
301 corresponded to the maximum time allowed to the subject to repeat the words they
302 remembered. No feedback was provided to participants, and their answers were recorded
303 in audio files.

304 The subject repeated this task 16 times in the five noise conditions. For each noise
305 condition, the cognitive load was either:

- 306 • high, by providing a list of words that was longer than the subject’s mnemonic
307 capacity. In our case, if the subject’s mnemonic value was n , the list displayed on the
308 screen during the task with a high cognitive load would have $n+2$ words.
- 309 • low, which corresponded to a list of words that was shorter than the subject’s
310 mnemonic capacity. In our case, if the subject’s short-term memory was n , the list
311 displayed on the screen during the task with a low cognitive load would have $n-1$
312 words.

313 The subject performed the memorisation task eight times with a high cognitive load and
314 eight times with a low cognitive load in each sound condition. The order of presentation of
315 each series (with a high cognitive load or low cognitive load) was random.

316 The lists of French words used for this task were chosen from the lexical lists published by
317 D. Dubois and J. Poitou (2002) [34]. The categories used during this experiment were:
318 Animals, Drinks, Buildings, Furniture and Sports. Each category consisted of 30 words,
319 selected based on their frequency of use in everyday language. One category was used for
320 one condition. The order of categories of words was different for each condition.

321 For each condition, one of the 16 9-minute sound signals mentioned in part 2.2 was used.
322 Permutation of these signals was performed on the principle of a (4x4) Latin square or
323 Klein group.

324 During the experiment, the subject was exposed to five different noise conditions, one of
325 which was the control condition, or silence.

326 During each condition and for each task, the sound environment was initiated during the
327 last second of the display of the fixation cross and was terminated during the last second of
328 the white screen, before the appearance of the “Over to you” message.

329 The order in which conditions were presented was different for each subject. This order
330 was based on a (5x5) Latin square to limit order effects linked to the sound source.

331 After completing the experiment in each condition, the experimenter entered the booth to
332 ask the participant if he had found the task difficult and where the difficulty was
333 encountered. This intervention also allowed the subject to take a rest between conditions.
334 This rest period ensured that the results for each condition reflected only the sound
335 annoyance. After this intervention, the subject answered three questionnaires in relation to
336 the condition that they had just completed. These questionnaires were presented to the
337 subject on a graphical interface in the following order:

- 338 • The first questionnaire assessed the sound annoyance simply by asking the question:
339 “Were you perturbed by the noise?” [35]. Answers were given on a Visual Analog
340 Scale (VAS).
- 341 • The second questionnaire, NASA-RTLX, assessed the mental workload through five
342 questions based on: mental demand, temporal demand, performance, effort and
343 frustration [28, 30].

344 • The third questionnaire was the Thayer questionnaire [33] which the subject had
345 already completed before starting the experiment.

346 For the control condition (silent background), the question on sound annoyance was,
347 naturally, not presented.

348

349 **3. Results**

350 Statistical analyses were performed using STATA software (version 14).

351 **3.1. Overall results**

352 During the mnemonic span test, participants had an average mnemonic value of four (1
353 participant had a mnemonic value of two (single element to be retained for the low
354 cognitive load); 11 participants had a mnemonic value of three; 19 participants had a
355 mnemonic value of four; 21 participants had a mnemonic value of five; 2 participants had a
356 mnemonic value of six; and 1 participant had a mnemonic value of seven). Therefore, the
357 numbers of items to be retained during the main phase therefore varied for each
358 participant between one and six for the low cognitive load condition and from 4 to 9 for the
359 high cognitive load condition.

360 *3.2 Decrease in performance*

361 For each sound condition, performance was calculated using the following relation [18, 30]:

$$362 \text{ Performance} = 100 * \left(\frac{\text{number}_{correct\ words} - \text{number}_{incorrect\ words}}{\text{number}_{total\ words}} \right) \quad (3)$$

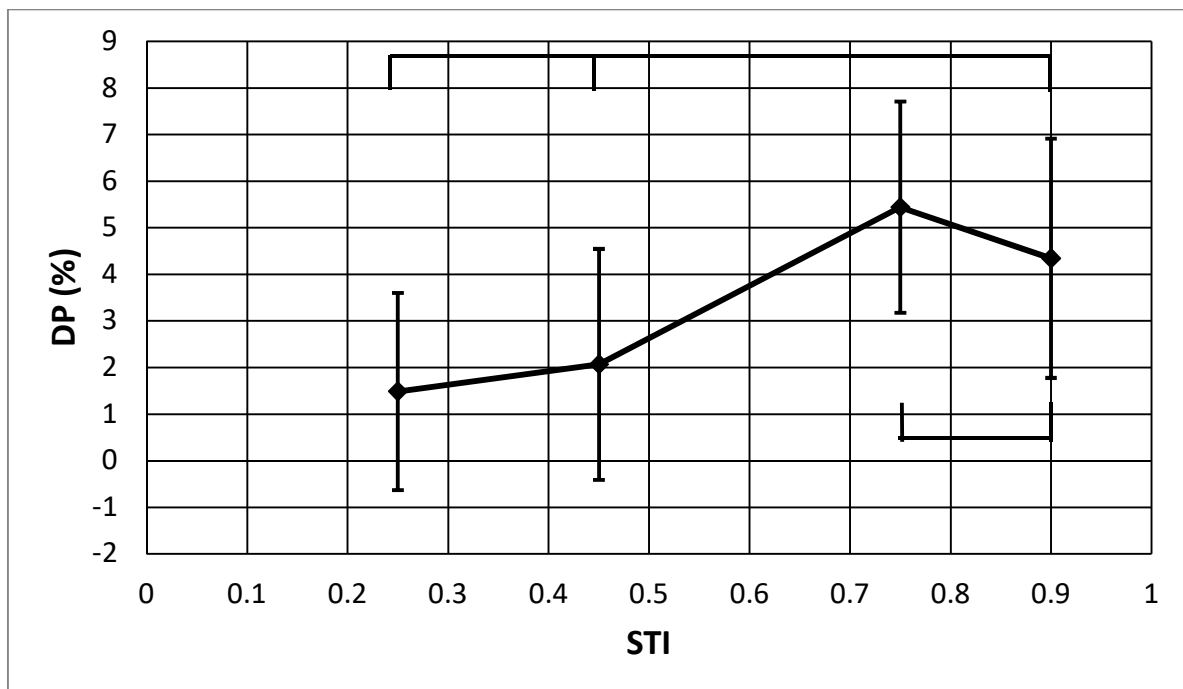
362

363 Then, relation (1) was used to determine the DP for each individual noise condition.

364 The effect of the running order of each condition was also analysed. For example, for the
365 silent condition, participants were split into five groups of 11 subjects based on the running

366 order (those who were presented with the silent condition first, those who were presented
367 with the silent condition second, etc.). The average performances for the five groups were
368 compared and revealed no effect of the running order for the silent condition. The same
369 procedure was applied for the four other conditions, and no effect of the running order was
370 found in any case. Subsequent analyses were therefore performed by grouping all
371 participants.

372 Analyses of variance (repeated measures) were performed, the dependent variables were a
373 parameter measured (performance, annoyance, mental workload) and the intrasubject
374 factor was the sound condition. For the decrease in performance (where the intrasubject
375 factor had four levels, corresponding to the four STI values), the effect of the STI was
376 significant **though small**: $F(3, 162) = 4.25, p < 0.01, \eta^2 = 0.015$. Figure 4 shows how the DP
377 changed with the increasing STI. The horizontal bars link the STI levels for homogeneous
378 groups as defined by Tukey's HSD.



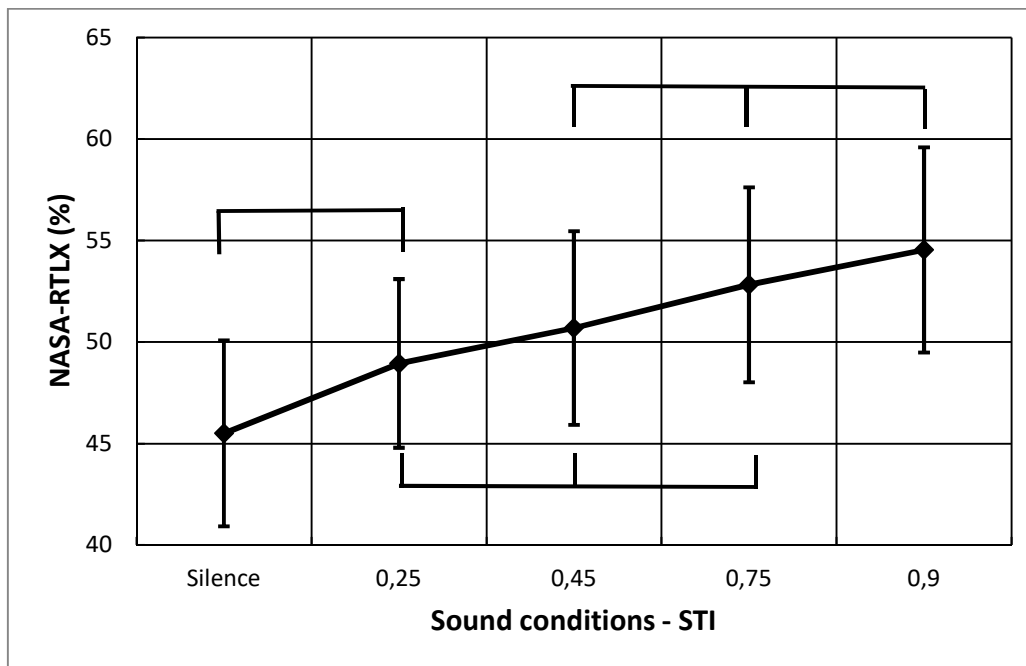
379

380 Figure 4. The DP with confidence intervals (95%) as a function of STI

381 Analysis of variance can also be performed for the DP measured in the two different
382 cognitive load conditions. In this case, two intrasubject factors exist: STI (4 levels) and
383 cognitive load (2 levels). The effect of cognitive load was not significant ($F(1,54) = 3.01$, $p =$
384 0.055), nor was there any interaction between cognitive load and STI values (data not
385 shown, $F(1,54) = 0.87$, $p = 0.46$). However, the significant effect of STI was still visible
386 (similar to that shown in figure 4). In contrast, when raw performance was considered, the
387 cognitive load was a significant factor ($F(1,54) = 752.7$, $p < 0.0001$, $\eta^2 = 0.93$). The average
388 performance, calculated across all sound conditions, was 92% in the low cognitive load
389 condition and 55% in the high cognitive load condition. As expected, the customization of
390 the task throughout the measurement of the mnemonic span value (n) did not influence the
391 ISE [36-38].

392 The sound condition also influenced the mental workload estimated by participants, as
393 indicated by their responses to the NASA-RTLX questionnaire. As shown in figure 5, this
394 workload was lower in the silent condition and increased constantly with the intelligibility
395 of speech ($F(4,216) = 7.5$, $p < 0.0001$, $\eta^2 = 0.034$).

396

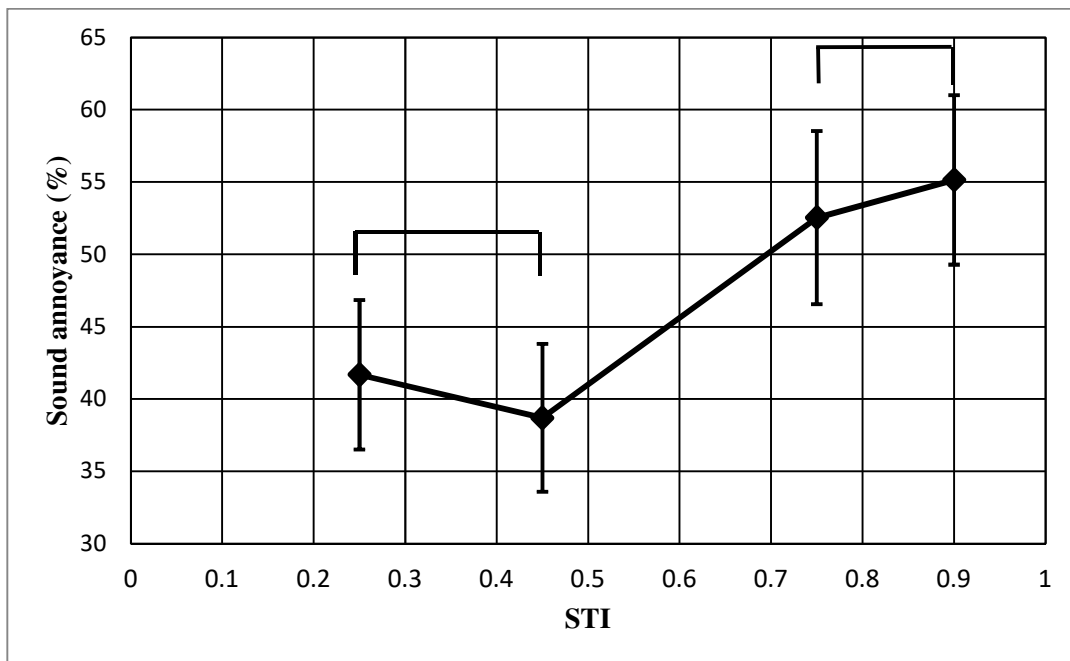


397

398 Figure 5. The NASA-RTLX score with confidence intervals (95%) as a function of sound
 399 conditions

400

401 A similar result was found for the sound annoyance described by participants, which was
 402 only assessed for the noisy situations. As shown in figure 6, the effect of the STI was
 403 significant ($F(3, 162) = 8.6, p < 0.001, \eta^2 = 0.051$) and the curve had a similar general profile
 404 to that linking STI and performance (figure 4).



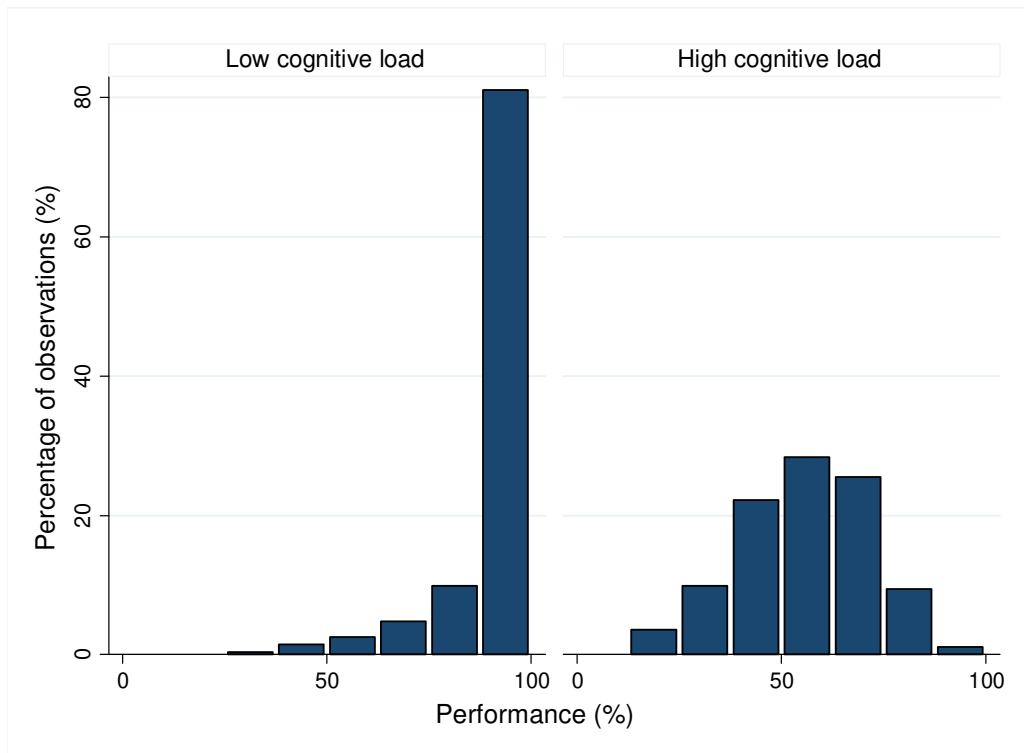
405

406 Figure 6. Sound annoyance with confidence intervals (95%) as a function of STI

407

408 **3.2. Supplementary findings**

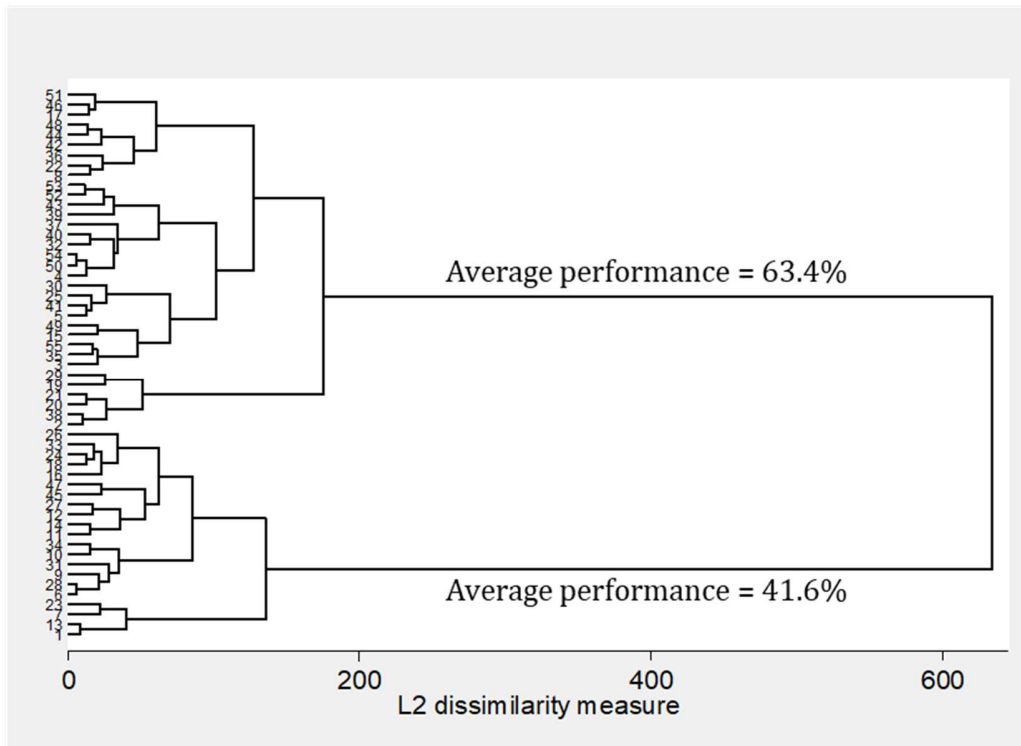
409 The results of the mnemonic span test performed at the start of the experiment, were used
 410 to personalise the complexity of the task. Analysis of data showed that for the low cognitive
 411 load condition the individual performances were relatively uniform: apart from four
 412 participants, the average individual performances in this situation were between 80 and
 413 100% (figure 7).



414
 415 Figure 7. Histograms of performance for the low cognitive load condition (left) and the high
 416 cognitive load condition (right)

417
 418 In contrast, in the high cognitive load condition, much greater interindividual variability
 419 was observed, extending from 26% to 85%.

420 The study of performance was continued by clustering the subjects with respect to their
 421 performances during the high cognitive load condition in the five sound conditions
 422 (hierarchical clustering, Euclidian distance and Ward method). This analysis clearly
 423 revealed two separate clusters of participants (figure 8), composed of 21 and 34 subjects.



424

425

Figure 8. Dendrogram for group cluster analysis using Ward's method

426

Observation of the average performances in the two groups, as a function of the sound

427

conditions (figure 9), confirmed that the two groups of subjects were separated based on

428

their general performance levels. The average performances for the two groups were 41.6%

429

(group 1) and 63.4% (group 2) in this high cognitive load condition. It should be noted that

430

the high performances observed for group 2 were also detectable in the low cognitive load

431

condition, where average performances were 86% for group 1 and 97% for group 2.

432

However, either the mental workload or the perceived noise annoyance did not depend on

433

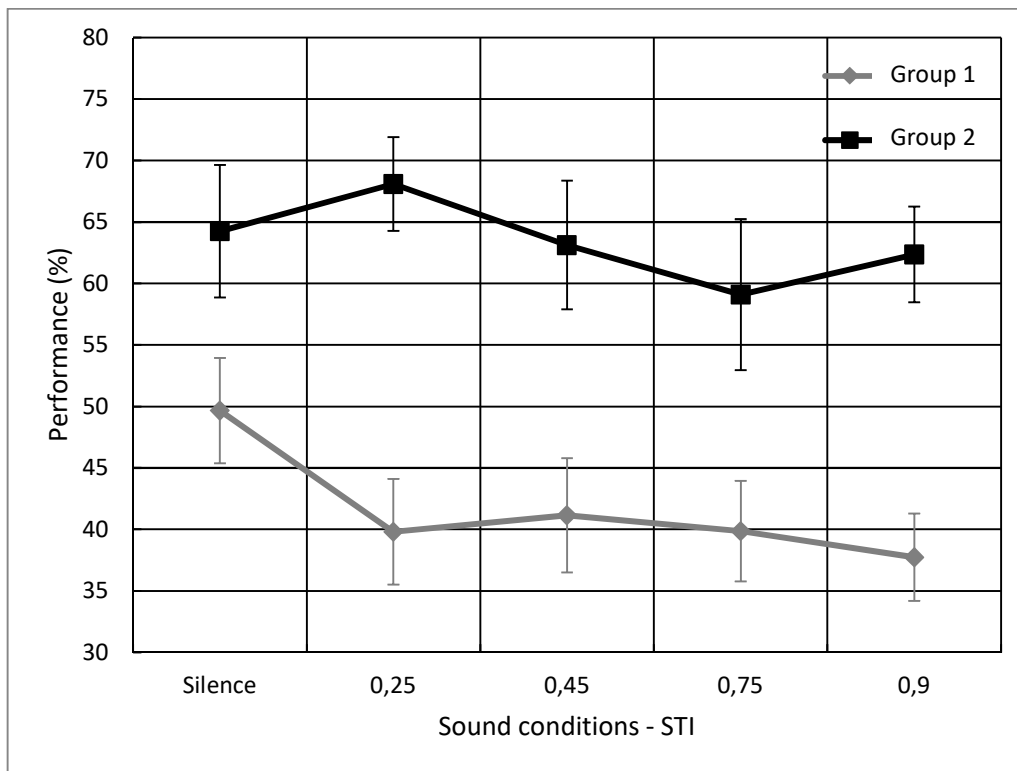
the group factor established by the cluster analysis (respectively $F(1,53) = 0.38, p = 0.5, \eta^2 =$

434

0.002 and $F(1,53) = 0.05, p = 0.82, \eta^2 = 0.001$). As expected, the short-term memory task

435

was similarly demanding for all participants.



436
 437 Figure 9. Performance confidence intervals (95%) during the short-term memory task as a
 438 function of sound conditions for the two groups of participants

439

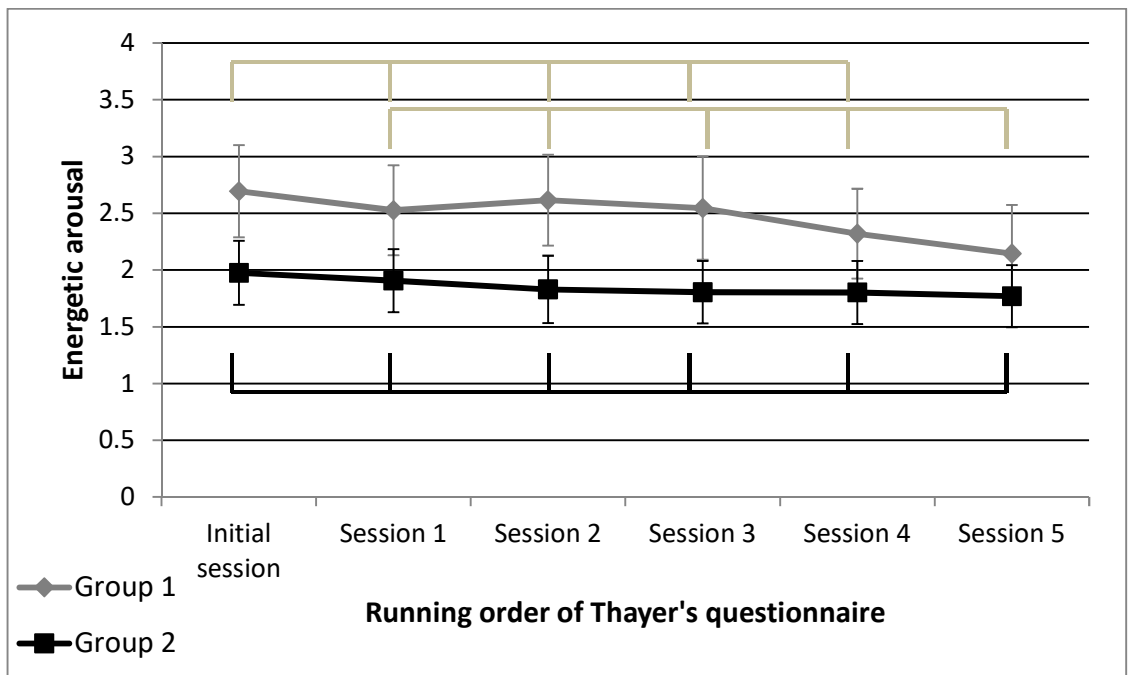
440 3.3. Energetic arousal

441 The final measurement performed during this study was that of energetic arousal of the
 442 participants at the beginning of and during the experiment. This measurement allowed us
 443 to identify any effect of “fatigue” that the participants might encounter during the
 444 experiment. Energetic arousal was measured using the Thayer questionnaire. This
 445 questionnaire was completed between each sound condition (six sessions where the first
 446 session is considered as the initial session). To verify that the time at which the experiment
 447 was started did not affect our results, an ANOVA was performed on results from the initial
 448 session performed before starting the short-term memory task. Participants were

449 separated into three groups depending on the time of day at which the experiment was
450 performed (approximately 10 AM, 2 PM and 4 PM). The analysis showed that there was no
451 significant difference between groups based on time of day ($F(2,53)=0.44$, $p=0.65$).
452 Similarly, the sound conditions did not significantly affect the energetic arousal
453 ($F(4,212)=0.76$, $p=0.55$).

454 In figure 10, the average level of the energetic arousal (which can vary between one and
455 four) is plotted as a function of the running order of the questionnaire throughout the
456 experiment, for the two groups of participants. The initial session was before the start of
457 the experiment, and session five was after the final sound condition. Energetic arousal can
458 be seen to decrease over time. A repeated-measure ANOVA revealed this effect of time on
459 the energetic arousal to be significant ($F(5,265)=4.43$, $p<0.001$, $\eta^2 = 0.016$). A post-hoc
460 Tukey HSD, represented by the horizontal bars in figure 8, showed that a significant
461 difference exists between the first test completed (initial session) and the last test (session
462 5).

463



464

465 Figure 10. Energetic arousal with confidence intervals (95%) as a function of the running
 466 order of Thayer's questionnaire

467 This analysis also revealed a significant difference between the two groups of participants
 468 ($F(1,53)=8.46, p<0.01, \eta^2 = 0.142$). In addition, using a post-hoc Tukey HSD (represented by
 469 the horizontal grey bars on figure 10), a significant decrease of the energetic arousal was
 470 detected during the experiment for group 1, with a significant difference between session
 471 five and the initial session. In contrast, the alertness of the second group, although lower
 472 overall, remained constant throughout the experiment.

473

474 4. Discussion

475 This study aimed to explore how high values of speech intelligibility affected the
 476 performance during a short-term memory task as well as subjective indicators of sound
 477 annoyance and mental workload. A method to adapt the main task to the mnemonic

478 capacities of the participants was proposed. The results of the experiment were analysed
479 using previously proposed indicators to better understand how speech intelligibility
480 affected the participants in this experiment.

481 **4.1. Main objectives**

482 **4.1.1. Decrease in performance**

483 For this type of experiment, DP is an indicator reflecting the impact of sound during a
484 complex task. The analysis of variance on the results clearly showed an effect of
485 intelligibility on the DP. This effect is relevant to previously published results for the
486 experiments described in part 2.1. As one of the aims of this study was to identify a specific
487 effect of high STI values, we examined values beyond those studied in most publications. As
488 shown in figure 3, no increase in DP was observed for STI values equal to 0.9. These results
489 therefore appear to support the Hongisto model (described by equation (2)) [15] rather
490 than the linear model.

491 **However, our results reveal that the choice of STI value = 0.5 to calculate a distraction**
492 **distance is not relevant, based on the effects of speech intelligibility on the DP.** The ISO
493 3382-3 standard (2012) [16] indicates that for an STI of less than 0.5, the DP decreases
494 rapidly. However, our results showed no significance between the different noise
495 conditions for STIs of 0.25, 0.45, and 0.9. The choice of this target value remains to be
496 consolidated.

497

498 **4.1.2. Mental workload**

499 The results of the statistical analysis show the same overall effect of STI on the mental
500 workload as that reported by Ebissou et al. (2015) [18]. This effect appears to be linear,

501 unlike the DP. However, a second objective of this study was to decrease interindividual
502 variability of the mental workload by adapting the task to the mnemonic capacity of each
503 participant. Ebissou et al. (2015) [18] and Brocolini et al. (2016) [30] showed that with
504 regard to a cluster analysis based on the individual performances during the short-term
505 memory task, a group of participants were better in performing the task than the rest of the
506 panel. This high-performers group had also a lower mental workload. As expected in this
507 study, this distinction of the mental workload between the two groups is not significant.
508 Even though the performances of the two groups were significantly different, the mental
509 workload required by the task was perceived in the same way. For all the participants of the
510 experiment the mental workload seems to increase linearly with the STI.

511

512

513 **4.1.3. Noise annoyance**

514 The results of sound annoyance presented here corroborate those of Schlittmeier et al.
515 (2008) [23] and Liebl et al. (2016) [22]. This indicator behaves in the same way across the
516 three studies. For STI values between 0.2 and 0.6, the sound annoyance was not
517 significantly different. When intelligibility was very high, results from the three studies
518 indicated a significant increase in sound annoyance.

519 As with the results of the NASA-RTLX, the sound annoyance was considered similar by the
520 two groups of participants. We can also report that the sound annoyance is sensitive to the
521 effect of speech intelligibility for all participants in our study.

522

523 **4.1.4. Performance**

524 Two groups of participants were formed based on their performance in the different sound
525 conditions. **In coherence** with the literature [27], belonging to one of the two groups is not
526 linked to mnemonic capacities of each group (Mann-Whitney U test, $z=0.65$, $p>0.5$).
527 Interestingly, we observed that the two groups presented a net mean age difference: 47.8-
528 years-old for the group performing less well, 37.8-years-old for the group performing
529 better. This difference was significant (Mann-Whitney U test, $Z = 3.867$, $p = 0.0001$). Other
530 studies have also reported an age effect on memory tasks [39, 40]. The discussed reason to
531 explain this effect is the decrease of the inhibition capacities, rather than the decrease of the
532 memorisation capacities per se. Thus, the participants in the first group (who had a higher
533 mean age) have “more difficulty” ignoring background noise during a short-term memory
534 task. In the present study, this lack of inhibition can be measured by examining the
535 sensitivity of the interference effect. Indeed, we can consider intelligible speech as an
536 irrelevant information which perturbs participants during a short-term memory task of
537 words. However, the results from the STROOP task, which measures this sensitivity to the
538 interference effect showed no significant difference between the two groups (Mann-
539 Whitney U test, $z = 0.78$ and $p = 0.43$) for the strong interference index.

540

541 **4.1.5. Energetic arousal**

542 The results relating to energetic arousal join those of performance when it comes to
543 separate participants into two groups. Thus, there is a significant difference between the
544 two groups of participants in terms of energetic arousal. However, the effect is opposite to
545 that which could have been expected since the most aroused participants were those who
546 performed least well (group 1). This group also appeared to be the most sensitive to the

547 decrease of their arousal over the course of the experiment. Their level of energetic arousal,
548 although higher, decreased significantly during the experiment, whereas it remained stable
549 for the best performing group. According to Galy et al. (2012) [41], the level of energetic
550 arousal determines the amount of cognitive resources that can be mobilised by individuals.
551 In this study, this arousal does not explain variations in performance for individuals as a
552 function of speech intelligibility. It therefore appears that the progressive decrease in
553 energetic arousal over the course of the experiment in individuals performing less well is a
554 reflection of the effect of a factor which was not considered here. This factor could be the
555 cognitive assessment of the task (e.g. study of the implementation of strategies employed
556 during the recall phase and the cognitive load that it represents). The group performing less
557 well which was composed of older subjects on average, may feel more “threatened” by the
558 task than the group composed of younger subjects. This hypothesis is in agreement with the
559 results obtained by Galy and Mélan (2015) [26] who reported that, during a mental
560 calculation task, individuals who considered the task “threatening” had lower performances
561 than others, even when their arousal levels were high. Thus, several studies report that
562 memorisation tasks are considered more “threatening” as age increases (ex.: Chasteen et al.
563 (2005) [42]).

564

565 **5. Conclusion**

566 Through this study, we examined the effect of speech intelligibility on performance in a
567 short-term memory task, using an STI range over which DP varies strongly (values of STI
568 between 0.2 and 0.7) and DP stays constant (high values of STI). Our results did not show
569 strong variations of DP between the two lower STI values. Therefore, the target values

570 presented in the ISO 3382-3 standard (2012) [16] could not be confirmed. In the high
571 intelligibility range the results fit better with the Hongisto predictive model [15] than with a
572 linear model. Participants were divided into two groups according to performance. This
573 separation could be related to the mean age difference in each group but the exact reason
574 still stays unclear.

575 This study was novel in customizing the short-term memory task with the mnemonic span
576 value for each participant. The customization did not reduce the interindividual variability
577 of performance, as it could be expected from the existing literature. On the other hand, the
578 analysis showed that the interindividual variability of the mental workload did reduce. This
579 approach could be used to measure the influence of finer variations of the sound
580 environment on mental workload during a short-term memory task.

581

582 **References**

- 583 1. AFNOR, *Acoustique – Performances acoustiques des espaces ouverts de bureaux*. 2016.
- 584 2. Boyce, P.R., *Users' assessments of a landscaped office*. Journal of Architectural Research,
585 1974. **3**: p. 44-62.
- 586 3. Klitzman, S. and J.M. Stellman, *The impact of the physical environment on the*
587 *psychological well-being of office workers*. Social Science & Medicine, 1989. **29**(6): p.
588 733-742.
- 589 4. Kaarlela-Tuomaala, A., et al., *Effects of acoustic environment on work in private office*
590 *rooms and open-plan offices – longitudinal study during relocation*. Ergonomics, 2009.
591 **52**(11): p. 1423-1444.
- 592 5. Sundstrom, E., et al., *Office Noise, Satisfaction, and Performance*. Environment and
593 Behavior, 1994. **26**(2): p. 195-222.
- 594 6. Pierrette, M., et al., *Noise effect on comfort in open-space offices: development of an*
595 *assessment questionnaire*. Ergonomics, 2015. **58**(1): p. 96-106.
- 596 7. Perrin Jegen, N. and P. Chevret, *Effect of noise on comfort in open-plan offices:*
597 *application of an assessment questionnaire*. Ergonomics, 2016: p. 1-12.
- 598 8. Salamé, P. and A. Baddeley, *Disruption of short-term memory by unattended speech:*
599 *Implications for the structure of working memory*. Journal of Verbal Learning and Verbal
600 Behavior, 1982. **21**(2): p. 150-164.

- 601 9. Salamé, P. and A. Baddeley, *Effects of background music on phonological short-term*
602 *memory*. The Quarterly Journal of Experimental Psychology Section A, 1989. **41**(1): p.
603 107-122.
- 604 10. Banbury, S. and D.C. Berry, *Disruption of office-related tasks by speech and office noise*.
605 British Journal of Psychology, 1998. **89**(3): p. 499-517.
- 606 11. LeCompte, D.C., C.B. Neely, and J.R. Wilson, *Irrelevant speech and irrelevant tones:*
607 *The relative importance of speech to the irrelevant speech effect*. Journal of Experimental
608 Psychology: Learning, Memory, and Cognition, 1997. **23**(2): p. 472-483.
- 609 12. Ellermeier, W. and K. Zimmer, *The psychoacoustics of the irrelevant sound effect*.
610 Acoustical Science and Technology, 2014. **35**(1): p. 10-16.
- 611 13. Steeneken, H.J.M. and T. Houtgast, *A physical method for measuring speech-*
612 *transmission quality*. The Journal of the Acoustical Society of America, 1980. **67**(1): p.
613 318-326.
- 614 14. IEC-60268-16, *Sound system equipment. Part 16: Objective rating of speech*
615 *intelligibility by speech transmission index*, I.E. Commission, Editor. 2011, International
616 Electrotechnical Commission: Switzerland.
- 617 15. Hongisto, V., *A model predicting the effect of speech of varying intelligibility on work*
618 *performance*. Indoor air, 2005. **15**(6): p. 458-468.
- 619 16. 3382-3, I., *Acoustics—Measurement of Room Acoustic Parameters. Part 3: Open Plan*
620 *Offices*. 2012, International Organization for Standardization Geneva.
- 621 17. Virjonen, P., J. Keränen, and V. Hongisto, *Determination of Acoustical Conditions in*
622 *Open-Plan Offices: Proposal for New Measurement Method and Target Values*. Acta
623 Acustica united with Acustica, 2009. **95**(2): p. 279-290.
- 624 18. Ebissou, A., E. Parizet, and P. Chevret, *Use of the Speech Transmission Index for the*
625 *assessment of sound annoyance in open-plan offices*. Applied Acoustics, 2015. **88**: p. 90-
626 95.
- 627 19. Ellermeier, W. and J. Hellbrück, *Is Level Irrelevant in 'Irrelevant Speech'? Effects of*
628 *Loudness, Signal-to-Noise Ratio, and Binaural Unmasking*. Journal of Experimental
629 Psychology. Human Perception & Performance, 1998. **24**(5): p. 1406-1414.
- 630 20. Haka, M., et al., *Performance effects and subjective disturbance of speech in acoustically*
631 *different office types – a laboratory experiment*. Indoor air, 2009. **19**(6): p. 454-467.
- 632 21. Jahncke, H., V. Hongisto, and P. Virjonen, *Cognitive performance during irrelevant*
633 *speech: Effects of speech intelligibility and office-task characteristics*. Applied Acoustics,
634 2013. **74**(3): p. 307-316.
- 635 22. Liebl, A., A. Assfalg, and S.J. Schlittmeier, *The effects of speech intelligibility and*
636 *temporal–spectral variability on performance and annoyance ratings*. Applied Acoustics,
637 2016. **110**: p. 170-175.
- 638 23. Schlittmeier, S.J., et al., *The impact of background speech varying in intelligibility:*
639 *Effects on cognitive performance and perceived disturbance*. Ergonomics, 2008. **51**(5): p.
640 719-736.
- 641 24. Schlittmeier, S.J., et al., *Algorithmic modeling of the irrelevant sound effect (ISE) by the*
642 *hearing sensation fluctuation strength*. Attention, Perception, & Psychophysics, 2012.
643 **74**(1): p. 194-203.
- 644 25. Zwicker, E. and H. Fastl, *Psychoacoustics: Facts and models*. Vol. 22. 2013: Springer
645 Science & Business Media.

- 646 26. Galy, E. and C. Mélan, *Effects of Cognitive Appraisal and Mental Workload Factors on*
647 *Performance in an Arithmetic Task*. Applied Psychophysiology and Biofeedback, 2015.
648 **40**(4): p. 313-325.
- 649 27. Sörqvist, P., J.E. Marsh, and A. Nöstl, *High working memory capacity does not always*
650 *attenuate distraction: Bayesian evidence in support of the null hypothesis*. Psychonomic
651 Bulletin & Review, 2013. **20**(5): p. 897-904.
- 652 28. Hart, S.G., *Nasa-Task Load Index (NASA-TLX); 20 Years Later*. Proceedings of the
653 Human Factors and Ergonomics Society Annual Meeting, 2006. **50**(9): p. 904-908.
- 654 29. Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of*
655 *empirical and theoretical research*. Advances in psychology, 1988. **52**: p. 139-183.
- 656 30. Brocolini, L., E. Parizet, and P. Chevret, *Effect of masking noise on cognitive*
657 *performance and annoyance in open plan offices*. Applied Acoustics, 2016. **114**: p. 44-55.
- 658 31. Wechsler, D., *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS—IV)*. 2014.
- 659 32. Bayard, S., J. Erkes, and C. Moroni, *Test du Stroop Victoria-Adaptation Francophone.*
660 *Matériel, consignes, procédure de cotation et données normatives*. 2009, CPCN-LR,
661 Gignac: Montpellier. p. 22.
- 662 33. Thayer, R.E., *Activation-deactivation adjective check list: current overview and structural*
663 *analysis*. Psychological Reports, 1986. **58**(2): p. 607-614.
- 664 34. Dubois, D. and J. Poitou, *Normes catégorielles pour vingt-deux catégories sémantiques*
665 *en français et dix catégories en allemand*. Cahiers du LCPE, 2002. **5**: p. 31-104.
- 666 35. Schlittmeier, S., et al., *The impact of background speech varying in intelligibility: Effects*
667 *on cognitive performance and perceived disturbance*. Ergonomics, 2008. **51**(5): p. 719-
668 736.
- 669 36. Beaman, C.P., *Auditory distraction from low-intensity noise: a review of the*
670 *consequences for learning and workplace environments*. Applied Cognitive Psychology,
671 2005. **19**(8): p. 1041-1064.
- 672 37. Ellermeier, W. and K. Zimmer, *Individual differences in susceptibility to the “irrelevant*
673 *speech effect”*. The Journal of the Acoustical Society of America, 1997. **102**(4): p. 2191-
674 2199.
- 675 38. Sörqvist, P., *High working memory capacity attenuates the deviation effect but not the*
676 *changing-state effect: Further support for the duplex-mechanism account of auditory*
677 *distraction*. Memory & Cognition, 2010. **38**(5): p. 651-658.
- 678 39. Bell, R., A. Buchner, and I. Mund, *Age-related differences in irrelevant-speech effects*.
679 Psychology and aging, 2008. **23**(2): p. 377.
- 680 40. Meijer, W.A., et al., *Verbal learning and aging: Combined effects of irrelevant speech,*
681 *interstimulus interval, and education*. The Journals of Gerontology Series B:
682 Psychological Sciences and Social Sciences, 2006. **61**(5): p. P285-P294.
- 683 41. Galy, E., M. Cariou, and C. Mélan, *What is the relationship between mental workload*
684 *factors and cognitive load types?* International Journal of Psychophysiology, 2012. **83**(3):
685 p. 269-275.
- 686 42. Chasteen, A.L., et al., *How feelings of stereotype threat influence older adults' memory*
687 *performance*. Experimental aging research, 2005. **31**(3): p. 235-260.
- 688