



HAL
open science

A new Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster and Comparative Analyses

Annie Moulin, Judith Vergne, Stéphane Gallégo, Christophe Micheyl

► To cite this version:

Annie Moulin, Judith Vergne, Stéphane Gallégo, Christophe Micheyl. A new Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster and Comparative Analyses. *Ear and Hearing*, 2019, 40 (4), pp.938-950. 10.1097/AUD.0000000000000675 . hal-02367082v2

HAL Id: hal-02367082

<https://hal.science/hal-02367082v2>

Submitted on 27 Aug 2023

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.

A new Speech, Spatial, and Qualities of Hearing Scale (SSQ)

Short-Form:

Factor, Cluster and Comparative Analyses

Annie Moulin^{1,2*}, Judith Vergne^{1,2}, Stéphane Gallego^{3,4,5}, Christophe Micheyl^{6,7,8}

¹. INSERM U1028, CNRS UMR 5292, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, F-69000 Lyon, France.

². Université C. Bernard, Lyon 1, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, University of Lyon, F-69 000 Lyon, France.

³. Institut des sciences et technologies de réadaptation (ISTR), University of Lyon, Lyon, France

⁴. Audition Conseil, France

⁵. Laboratory of Integrative and Adaptive Neurosciences (LNIA) UMR 7260, Aix-Marseille University-CNRS, Marseille, France.

⁶. Starkey Hearing Technologies, Créteil, France

⁷. Starkey Hearing Research Center, Berkeley, CA, USA

⁸. INSERM U1028, CNRS UMR 5292, Cognition and Auditory Perception Team (CAP), Lyon Neuroscience Research Center, University of Lyon, F-69000 Lyon, France.

*Corresponding author: annie.moulin@cnrs.fr

This is the final version of the manuscript accepted for publication in “*Ear and Hearing*”.
ISSN: 0196-0202, ESN: 1538-4667.

The supplementary material (xlsx file) is available at: <https://hal.science/hal-02367082v1>

The publisher’s version is available at:

https://journals.lww.com/ear-hearing/abstract/2019/07000/a_new_speech_spatial_and_qualities_of_hearing.16.aspx

Full reference of the final version:

Moulin, A., Vergne, J., Gallego, S., & Micheyl, C. (2019). A New Speech, Spatial, and Qualities of Hearing Scale Short-Form : Factor, Cluster, and Comparative Analyses. *Ear and Hearing*, 40(4), 938-950. <https://doi.org/10.1097/AUD.0000000000000675>

1 **A new Speech, Spatial, and Qualities of Hearing Scale (SSQ) Short-Form:**
2 **Factor, Cluster, and Comparative Analyses**

3

4 Annie Moulin^{1,2*}, Judith Vergne^{1,2}, Stéphane Gallego^{3,4,5}, Christophe Michey^{6,7,8}

5

6 ^{1.} INSERM U1028, CNRS UMR 5292, Lyon Neuroscience Research Center, Brain Dynamics
7 and Cognition Team, F-69000 Lyon, France.

8

9 ^{2.} Université C. Bernard, Lyon 1, Lyon Neuroscience Research Center, Brain Dynamics and
10 Cognition Team, University of Lyon, F-69 000 Lyon, France.

11 ^{3.} Institut des sciences et technologies de réadaptation (ISTR), University of Lyon, Lyon,
12 France

13 ^{4.} Audition Conseil, France

14 ^{5.} Laboratory of Integrative and Adaptive Neurosciences (LNIA) UMR 7260, Aix-Marseille
15 University-CNRS, Marseille, France.

16 ^{6.} Starkey Hearing Technologies, Créteil, France

17 ^{7.} Starkey Hearing Research Center, Berkeley, CA, USA

18 ^{8.} INSERM U1028, CNRS UMR 5292, Cognition and Auditory Perception Team (CAP),
19 Lyon Neuroscience Research Center, University of Lyon, F-69000 Lyon, France.

20

21 *Corresponding author: annie.moulin@cns.fr

22 Postal address : DYCOG - Dynamique cérébrale et Cognition, Centre de Recherche en
23 Neurosciences de Lyon, CRNL, Inserm U1028 - CNRS UMR5292, CH Le Vinatier, Bâtiment
24 452, 95 Bd Pinel, 69675 BRON Cedex France.

25

26 **Conflicts of interest and sources of funding**

27 Author CM is supported by Starkey Hearing Technologies, a private entity and manufacturer
28 of hearing technology. Author SG is supported by Audition Conseil, a private company and
29 group of audiology clinics. Other than through funding of these two coauthors' salaries, the
30 sponsors for this study had no involvement in the design of the study, the data analysis, or the
31 writing of the manuscript.

32 This work was supported in part by the "Auvergne-Rhône-Alpes" region (research program
33 "Effecbruit"); the "Fondation de l'Avenir" and "Visaudio" (research program ET4-738-VI4-
34 001); by the LABEX CELYA (ANR-11-LABX-0060) of Université de Lyon, France and the
35 LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program
36 "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research
37 Agency (ANR).

38

39 **A new Speech, Spatial, and Qualities of Hearing Scale (SSQ) Short-Form:**
40 **Factor, Cluster, and Comparative Analyses**

41

42 Annie Moulin, Judith Vergne, Stéphane Gallego, Christophe Michey

43

44 **Acronyms and abbreviations**

45 **AUC:** Area under the curve (for Receiver Operating Characteristics curves)

46 **SSQ5s:** 5-item SSQ short-form, obtained as a subset of the SSQ

47 **SSQ12s:** 12-item SSQ short-form, obtained as a subset of the SSQ

48 **SSQ15s:** 15-item SSQ short-form, obtained as a subset of the SSQ

49 **15iSSQ:** new 15-item SSQ short-form

50 **15iSSQs:** new 15-item SSQ short-form, obtained as a subset of the SSQ

51 **SSQ:** Speech Spatial and Qualities of hearing scale

52 **Speech:** Speech subscale of the SSQ

53 **Spatial:** Spatial-hearing subscale of the SSQ

54 **Qualities:** Qualities of hearing subscale of the SSQ

55 **HI:** Hearing-impaired

56 **HIHA:** Hearing-impaired, hearing-aid wearers

57 **HIN:** Hearing-impaired, non-hearing-aid wearers

58 **NH:** Normal-hearing subjects

59 **PTA:** Pure-tone threshold average (0.5, 1, 2 and 4 kHz)

60 **ROC:** Receiver Operating Characteristics curves

61 **SD:** standard deviation

62 **NA:** Non-applicable answers

63 **F-ANOVA:** Friedman's analysis of variance

64 **R-ANOVA:** repeated-measures analysis of variance

65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88

ABSTRACT

Objectives: The objective of this work was to build a 15-item short-form of the Speech Spatial and Qualities of Hearing Scale (SSQ) that maintains the three-factor structure of the full form, using a data-driven approach consistent with internationally recognized procedures for short-form building. This included the validation of the new short-form on an independent sample and an in-depth, comparative analysis of all existing, full and short SSQ forms.

Design: Data from a previous study involving 98 normal-hearing (NH) individuals and 196 hearing-impaired, non-hearing-aid wearers (HIN), along with results from several other published SSQ studies were used for developing the short-form. Data from a new and independent sample of 35 NH and 88 HI hearing-aid wearers (HIHA) were used to validate the new short-form. Factor and hierarchical cluster analyses were used to check the factor structure and internal consistency of the new short-form. In addition, the new short-form was compared to all other SSQ forms, including the full SSQ, the German SSQ15, the SSQ12, and the SSQ5. Construct validity was further assessed by testing statistical relationships between scores and audiometric factors, including pure-tone threshold averages (PTAs) and left/right PTA asymmetry. Receiver-operating characteristic (ROC) analyses were used to compare the ability of different SSQ forms to discriminate between NH and HI (HIN and HIHA) individuals.

Results: Compared all other SSQ forms, including the full SSQ, the new short-form showed negligible cross-loading across the three main subscales, and greater discriminatory power between NH and HI subjects (as indicated by a larger area under the ROC curve), as well as between the main subscales (especially Speech and Qualities). Moreover, the new, 5-item Spatial subscale showed increased sensitivity to left/right PTA asymmetry. Very good internal

89 consistency and homogeneity, and high correlations with the SSQ were obtained for all short
90 forms.

91

92 **Conclusion:** While maintaining the three-factor structure of the full SSQ, and exceeding
93 the latter in terms of construct validity, and sensitivity to audiometric variables, the new 15-
94 item SSQ, the 15iSSQ, affords a substantial reduction in the number of items and thus, in test
95 time. Based on overall scores, Speech subscores, or Spatial subscores, but not Qualities
96 subscores, the 15iSSQ appears to be more sensitive to differences in self-evaluated hearing
97 abilities between NH and HI subjects than the full SSQ.

98

99

100 **Key words:** Self-report Measure; Short form building; Hearing Disability; Hearing Loss;
101 Spatial Hearing; Speech.

102

103

INTRODUCTION

104

105

106

107

108

109

110

111

Self-report outcome measures have become an essential component in the evaluation of rehabilitation benefits for patients. In the field of audiology, several scales have been developed for the self-evaluation of various aspects of hearing, such as speech perception, binaural hearing, or hearing in challenging situations. Among these, the Speech Spatial and Qualities of Hearing scale (SSQ) designed by Gatehouse and Noble (2004) is a widely used self-report measure of hearing, and has already been translated into several languages (for example, Dutch (Demeester et al. 2012), Korean (Kim et al. 2017), German (Kiessling et al. 2011), French and Portuguese (Gonzalez et al. 2015).

112

113

114

115

116

117

118

119

120

121

122

In its full form, the SSQ includes 49 items. One advantage of this relatively large number of items, is that it makes it possible to explore some very specific aspects of hearing via three main subscales (Speech, Spatial and Qualities) and 10 pragmatic subscales (Gatehouse & Akeroyd 2006). One disadvantage relates to the substantial amount of time and effort required for completion. Using a French version of the SSQ, Moulin et al. (2015) found that the self-assessed time to complete the scale in a hearing-impaired (HI) group varied from 10 minutes to one hour, with more than 25% of respondents reporting completion times above 25 minutes. This makes it difficult to use the SSQ in routine clinical use, or for swift assessments of hearing. Moreover, the number of missing responses tended to increase over the last third of the scale, with more missing responses for the most difficult-to-read items (Moulin et al., 2015). Indeed, with over a thousand words, the SSQ can be quite taxing for respondents.

123

124

125

126

127

In order to address this problem, short-forms of the SSQ have been developed. Demeester et al. (2012) created a short-form with five items (SSQ5) out of the Dutch version of the SSQ, specifically for screening purposes. Using the UK version of the SSQ, Noble et al. (2013) created a short-form with 12 items (SSQ12) based on a large multi-center dataset. Neither of these short forms include the main three subscales. Indeed, Noble et al. (2013)'s aim

128 was “to compile a set of items that represent the scale as a whole”, and that reflect the 10
129 pragmatic subscales of the SSQ, as defined by Gatehouse and Akeroyd (2006). Validating a
130 German version of the SSQ, Kiessling et al. (2011) proposed a short-form containing 15 items,
131 with five items per main subscale. The resulting short-form has the potential to retain the
132 information present in the three main subscales of the SSQ. However, each of these short forms
133 was constructed based on results from a single data sample, an approach which should be
134 avoided if possible, because it tends to overestimate the expected reliability of the short-form
135 (Putnam & Rothbart, 2006), and fails to take into account potential variations of the optimal
136 subset of items across different data samples (Widaman et al. 2011). To our knowledge, except
137 for the SSQ5, which was built specifically for screening purposes, none of these short-forms
138 has been validated yet on a different sample of participants from the one used to build it, which
139 is one of the last recommended steps in short-form building. Indeed, the difficulty of short-form
140 building is the trade-off between the potential gain in time to administer the scale and the
141 potential loss in psychometric properties due to items reduction. This trade-off renders an
142 independent assessment of the reliability and validity of the new measure essential (Smith et al.
143 2000).

144 In this context, the main goals of the present study were as follows. Firstly, the aim was
145 to create a short-form of the SSQ using a data-driven approach, taking into account SSQ results
146 from as many studies as available, provided that the results were given per item, and following
147 the recommended guidelines for short form building (Smith et al, 2000; Widaman et al. 2011;
148 Stanton et al. 2002). Secondly, this work allowed us to evaluate the performances of the pre-
149 existing and newly created short-forms on a new, independent sample of participants. To this
150 aim, we tried to follow the recommendations of Smith et al. (2000) for avoiding the “sins of
151 short-form development”. Hence, the short form was built using a data-driven approach and
152 strategies specific to short-form building (Stanton et al. 2002), using published data from five

153 SSQ studies involving four different languages (Dutch, English, French, and German). First,
154 responses to items of the newly built short-form were extracted from existing data from a
155 previous study in which the full SSQ was administered to 196 hearing-impaired (HI)
156 participants (non-wearers of hearing-aids: HIN) and 100 normal-hearing (NH) participants
157 (Moulin et al., 2015). Subsequently, both the full SSQ and the newly built short-form were
158 administered to a new sample of 88 HI subjects (all hearing-aid wearers: HIHA) and 35 NH
159 subjects. In-depth analyses of all existing, full and short SSQ forms were performed. In
160 particular, factor and hierarchical cluster analyses were used to check the factor structure and
161 internal consistency of the new short-form. In addition, the new short-form was compared to
162 all other SSQ forms, including the full SSQ, the German SSQ15, the SSQ12, and the SSQ5.
163 Construct validity refers to how well the new scales cover the same content as the original full
164 scale and relates to external factors the same way as the original scale. With reliability, it is the
165 most important factor to respect when creating a short-form (John and Soto, 2009). This was
166 assessed by analyzing statistical relationships between scores and patients' characteristics, such
167 as age and audiometric factors, including pure-tone threshold averages (PTAs) and left/right
168 PTA asymmetry. In addition, the ability of the SSQ and all the short-forms to discriminate
169 between NH and HI (HIN and HIHA) individuals was compared using receiver-operating
170 characteristic (ROC) analyses. Those analyses consist of building curves that show the
171 sensitivity (ability to detect impairment) as a function of the specificity (proportion of correctly
172 recognized NH), as the cutoff score of the SSQ forms is moved across its range. This gives a
173 summary statistic (the area under the curve) allowing to compare the quality of the
174 discrimination between impairment and non-impairment.

MATERIAL AND METHODS

175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199

1. Short-form creation

In order to create the new short-form, we considered the following sources of information:

- Missing-response rates across items for the German SSQ (Kiessling et al. 2011), the English SSQ (Akeroyd et al. 2014), and the French SSQ (Moulin et al. 2015; Moulin & Richard 2016b).

- Results and normative values obtained per item (mean, standard deviations) on young adult NH subjects for the Dutch SSQ (Demeester et al. 2012), the English SSQ (Banh et al. 2012) and the French SSQ (Moulin et al. 2015).

- Factor analyses of the English SSQ (Akeroyd et al. 2014) and French SSQ (Moulin et al. 2015).

One of the recommended methods suggested by Widaman et al. (2011) for building short-forms, is to identify a subset of items in the full-length questionnaire, which maintains the factorial integrity of the main scale. For the SSQ, factor analyses have shown that the main factors correspond to the three main subscales, Speech, Spatial and Qualities (Akeroyd et al. 2014; Moulin et al. 2015). The recommended minimum number of items per subscale ranges between three and five (Loewenthal 2001). We opted to create a short-form with five items per subscale (15 items in total), thus keeping a possibility of further scaling down the questionnaire to three or four items per subscale. To distinguish this new short-form from other SSQ short-forms, namely, Noble et al.'s (2013) SSQ12 and Kiessling et al.'s (2011) SSQ15, we hereafter refer to it as the "15-item SSQ" (15iSSQ).

The selection of items for inclusion in the short-form used the three categories of "item quality" proposed by Stanton et al. (2002): (1) *Internal items qualities*, which refer to scale properties such as inter-item and total-to-item correlations, internal-consistency measures (e.g.

200 Cronbach's alpha), factor-analysis results, and subscale internal structure; (2) *External item*
201 *qualities*, which refer to the way the items interact with external factors (construct validity),
202 such as correlations between scores and hearing loss, ear asymmetry, age, or number of years
203 of education. (3) *Judgmental item qualities*, which refers to the subjective evaluation by the
204 scale user, of the relevance and adaptation to the patient's level of understanding. Indeed, some
205 items of the SSQ are quite long to read and can be difficult to understand, as shown by
206 readability scores, that range from easy to extremely difficult for some items of the quality
207 subscale (Moulin et al., 2015). Because these indices can yield contradictory information
208 regarding the "quality" of an item, item selection often reflects a compromise between different
209 evaluation criteria. To deal with this, Stanton et al. (2002) proposed to code the quality indices
210 for each item and to sort the items by their level of quality, starting from *external items quality*
211 such as item-level validity, followed by *judgmental quality and face validity*, and finally,
212 *internal-consistency item qualities*. As we wanted to eliminate, first, the most unreliable items,
213 we adapted this approach by devising a system of "penalty points", awarded for each item
214 according to the extent of its departure from several numeric criteria related to the quality
215 categories. The scale for "penalty points" was 0, 1, and 2; a penalty of 2 on one criterion meant
216 that the corresponding item was eliminated. The details of the penalty points awarded to each
217 item are provided in the supplemental table ST1. In a first phase, the following criteria were
218 used:

219 1.1 Elimination of items with the greatest percentage of missing answers and NA
220 responses (Missing >10% in English, French and German SSQ). Indeed, the number of missing
221 answers (subjects who ticked the "non applicable" box) can be quite high for some items, which
222 happen to be the same items across the different data samples (See fig. 1 showing comparison
223 of SSQ missing answers in Moulin et al. 2015). For shorter forms, this can be a problem, if
224 some items are systematically missing (Putnam & Rothbart, 2009). Therefore, the rate of

225 missing items was one of the criterion of choice for reducing the number of items for the short-
226 form.

227 1.2 Elimination of items with less than 20% of variance explained by a model
228 involving hearing loss, ear asymmetry, age, gender, and number of years of education (based
229 on an analysis of French SSQ data from Moulin et al. 2016). This allowed us to exclude
230 individual items that showed low correlations with hearing loss and ear asymmetry. Hence, this
231 criterion was used to reinforce the external validity of the scale, to ensure good correlation with
232 external factors such as hearing loss and hearing asymmetry.

233 1.3 Elimination of all items with the lowest communalities (<50%) in factor
234 analyses of the French SSQ and the English SSQ.

235 1.4 Elimination of items having a main-factor load lower than 0.60 and/or cross-
236 loadings larger than 0.20 (based on factor analyses of the French SSQ and the English SSQ).
237 The 1.3 and 1.4 criteria allow to ensure the maintenance of the internal structure of the SSQ,
238 with three well-defined subscales, in the short-form.

239 1.5 Elimination of items whose cutoff scores (defined as the mean – 2 SD), in
240 young NH subjects, were below 3.5. This was based on 98 subjects for the French SSQ, 103
241 subjects for the Dutch SSQ (Demeester et al. 2012), and 48 subjects for the English SSQ (Banh
242 et al. 2012). As the mean (and SD) across items correlate significantly between languages and
243 data samples (Moulin et al., 2015), this criterion allowed to eliminate items with too low a value
244 (defined as a value <3.5) in NH subjects. Indeed, the SSQ items that are low in NH subjects
245 tend to decrease the contrast between NH and HI subjects.

246 This first selection phase left a total of 23 items remaining, with 8 from Speech,
247 10 from Spatial and 5 from Qualities.

248 In a second phase, the criteria listed above were used more coarsely, in that the “penalty
249 points” were summed together rather than considered separately; in addition; four other criteria
250 were added:

251 2.1 Favoring items with scores not significantly predicted by unwanted factors,
252 such as the number of years of education or gender. Indeed, the scores of some items of the
253 spatial and quality subscales showed a significant correlation (albeit minor) with the number of
254 years of education of the patients (Moulin & Richard, 2016). Those items were amongst the
255 longest and most difficult to read by the patients. All things equal, eliminating preferably those
256 items from the short-form is likely to increase its validity.

257 2.2 Favoring, for the spatial scale, items showing a strong prediction by left/right
258 asymmetry in pure-tone thresholds, and eliminating items predicted strongly by ear asymmetry
259 for the other subscales. Indeed, as the spatial scale is strongly related to localization in space
260 and left/right asymmetry, favoring items correlating the most with ear asymmetry is likely to
261 reinforce the specificity of the spatial subscale.

262 2.3 Favoring, whenever possible, items already present in the SSQ12, the
263 SSQ15, or the SSQ5.

264 2.4. Favoring items with high item-to-total correlations. This criterion allows to
265 increase the reliability of the short-form, but as all SSQ items show a good item-to-total
266 correlation, the expected reliability of the short-form is high, regardless of item choice. Hence,
267 this was not our first criteria.

268 This second phase led to a 15-item short-form (#1.1, #1.4, #1.5, #1.6, #1.11, #2.2, #2.6,
269 #2.7, #2.11, #2.17, #3.4, #3.5, #3.6, #3.8, #3.9) having one item in common with the SSQ5,
270 five items in common with the SSQ12, and 7 items in common with the SSQ15. As an aside,
271 the SSQ12 and SSQ15 have five items in common, while the SSQ5 and SSQ12 have three items

272 in common (see supplemental table ST1). Importantly, the 15-item short-form thus created
273 conforms to the three main subscales of the SSQ.

274

275

276 **2. Short-form validation**

277 The short-forms were validated with two samples of subjects. Sample A contained 98
278 young NH subjects (aged 18 to 27 years; mean age = 20.8 years; SD = 2.2 years) and 196 HI
279 subjects (aged 18 to 88 years; age = 53.4 years; SD = 17.1 years), who did not use hearing-aids
280 (HIN). For the latter, the four-frequency (0.5, 1, 2, and 4 kHz) PTA was equal to 27.2 dB HL
281 (SD = 14.4 dB) on the best ear, and 41 dB HL (SD = 24 dB) on the worse ear, on average (Fig.
282 1). The HIN subjects formed a subset of an original sample of 216 HI subjects, which was used
283 in two previous publications, along with the 98 NH subjects (Moulin & Richard 2016a; Moulin
284 & Richard 2016b).

285 Sample B was a new sample, completely independent from the first one. It comprised
286 35 NH subjects aged 18 to 30 years (mean = 21.6 years, SD = 2.7 years), and 88 HI hearing-
287 aid users (HIHA) aged 19 to 94 years (mean = 73.6 years, SD = 16.2 years). HI participants
288 without HAs (sample A), as well as with HAs (sample B), were tested so as to document the
289 applicability of the F-SSQ to these two types of HI participants, the F-SSQ characteristics
290 having been only described in HI patients (and NH subjects) so far. The NH subjects were
291 recruited mostly from a pool of university undergraduates. They were free from any medical
292 condition, history of otologic pathology, medication, and heavy noise exposure. They
293 underwent pure-tone audiometry in half-octave steps from 125 to 8000 Hz using an
294 Interacoustic AC40 audiometer in a sound-proof booth. The HIHA subjects were recruited from
295 an audiology clinic. Their unaided pure-tone thresholds were measured at 250, 500 Hz, then at
296 half-octave steps from 500 to 8000 Hz, using an Aurical Astera. Aided thresholds were

297 measured using warble tones at these same frequencies, through speakers. Those patients had
 298 been wearing their hearing-aids for an average of 3.4 years (SD = 3.7 years) and an average of
 299 10.4 hours per day (SD = 5.1 hours). The average PTA for the best ear was 43.9 dB HL (SD =
 300 13.4 dB) and 53.8 dB HL (SD = 19.5 dB) for the worse ear (Fig. 1).

301 Subjects from sample B were administered the validated full French language version
 302 SSQ (Moulin et al., 2015) and the new short-form version at two different time points, separated
 303 by at least two hours and up to a few days, in randomized testing order; the short-form was
 304 presented first to half of the subjects and second to the other half of the subjects. The HIHA
 305 subjects were instructed to fill the SSQ as if they were listening through their hearing-aids. For
 306 all samples, the SSQ and the short-form were completed independently by the participants, after
 307 these were instructed by an audiologist; the interview mode was not used.

308 The research was conducted in agreement with the World Medical Association
 309 declaration of Helsinki and in agreement with the French law pertaining to biomedical research
 310 (Agreement number A-11-385, 'CPP Sud-Est IV').

311

312 **3. Data and statistical analyses**

313 ***3.1. Score calculations***

314 PTA was computed as the average hearing threshold (in dB HL) across 0.5, 1, 2, and 4
 315 kHz. Ear asymmetry was calculated as the difference in PTAs (in dB) between the left and right
 316 ears. SSQ scores were calculated over the full scale and over the different short-forms that were
 317 available in the literature: the SSQ5 (#1.8, #2.3 and #2.9, #3.9 and #3.14; its weighted version
 318 ($SSQ5w = [(1.8 \times 0.804) + (2.3 \times 0.770) + (2.9 \times 0.676) + (3.9 \times 0.806) + (3.14 \times 0.646)]$
 319 $/ 5$) (Demeester et al. 2012)), the SSQ12 (#1.1, #1.4 and #1.10 to #1.12; #2.6, #2.9 and #2.13;
 320 #3.2, #3.7, #3.9 and #3.14; (Noble et al. 2013) and the SSQ15 (#1.4, #1.5, #1.7, #1.9, #1.10;
 321 #2.5 to #2.7, #2.9 and #2.12; #3.3, #3.4) (Kiessling et al. 2011). Scores corresponding to the

322 new 15-item short-form, 15iSSQ, (#1.1, #1.4, #1.5, #1.6, #1.11, #2.2, #2.6, #2.7, #2.11, #2.17,
323 #3.4, #3.5, #3.6, #3.8, #3.9) were also calculated.

324 Scores were computed individually for every NH and HI subjects in samples A and B.
325 For sample B, two scores were computed for the 15iSSQ: one score was obtained by scoring a
326 subset of items from the SSQ questionnaire, hereafter referred to as the 15iSSQs, for ‘15-item
327 SSQ subset’; the second score was obtained by scoring the actual 15-item short-form, which
328 was administered separately to the subjects rather than as a part of the full SSQ. All the other
329 short-forms were obtained as a subset of the SSQ for both samples. Hence, to avoid confusion,
330 those other short-forms will be labelled SSQ15s, SSQ12s, SSQ5s.

331 For the SSQ15s and the 15iSSQ, both of which were designed to maintain the three-
332 subscale structure of the SSQ (Speech, Spatial *and* Qualities), three “differential” sub-scores
333 were calculated by subtracting sub-scores for the three subscales, pairwise (Qualities – Speech;
334 Qualities – Spatial and Speech – Spatial).

335 Similarly to Demeester et al. (2012), we calculated the SSQ-disability cut-offs, defined
336 as the mean scores obtained in normal hearing population, minus 2 standard deviations, for the
337 total scores obtained from SSQ and all short-forms, and the Speech, Spatial and Qualities
338 subscales.

339 **3.2 Internal-structure analysis**

340 Data from the various short-forms were analyzed with common factor analysis, using
341 the same parameters as for the full SSQ (see Moulin et al. 2015). Briefly, these analyses were
342 performed using the R package “Psych”, with the following parameters: correlation matrix of
343 15x15 (SSQ15i), maximum likelihood method for factors extraction, parallel analysis for
344 number of factors extraction, and oblique (oblimin) factor rotation.

345 Secondly, to determine whether factors of the short-forms reflect the same structure as
346 the full scale, hierarchical cluster analyses (Revelle’s ICLUST algorithm, (Revelle 1978;

347 Revelle 1979) were performed. Revelle's method is designed specifically to visualize
348 questionnaire scales and subscales. It relies on two indices: the alpha coefficient (mean split-
349 half reliability), a measure of internal consistency, and Revelle's beta coefficient (Zinbarg et al.
350 2005), a measure of factorial homogeneity; specifically, Revelle's beta coefficient is the worst
351 (i.e., the lowest) split-half reliability of a scale, and hence is lower than alpha. In short, the scale
352 structure is built starting from two item clusters that are most similar to each other; an item is
353 added to the initial two item clusters only if this addition improves the internal consistency
354 (measured by alpha) and/or the factorial homogeneity (measured by beta) of the cluster. The
355 results are shown using a hierarchical tree diagram of clusters that displays the internal sub-
356 structure of the scale, allowing the definition of homogeneous subscales. The tree diagram
357 connects increasingly less similar items and/or clusters from left to right: the most similar items
358 are combined first, and the most dissimilar items are added last. Alpha and beta coefficients are
359 provided for each cluster, and correlations coefficients are given for each connection between
360 clusters and/or items. An alpha above 0.8 and a beta above 0.7, with a difference between these
361 two coefficients lower than 0.1, are indicative of good homogeneity and good internal
362 consistency (Cooksey & Soutar 2006). A goodness-of-fit index was used to compare the quality
363 of the structures. Specifically, the fits of the different models to the data were assessed using
364 root-mean-square residuals (RMSRs). An RMSR lower than 0.05 indicates a good fit; RMSRs
365 comprised between 0.05 and 0.08 indicate a fair fit (Cooksey & Soutar 2006; Fabrigar et al.
366 1999). This technique is complementary to the more classical approach of factor analysis and
367 is less method-dependent (Cooksey & Soutar 2006). The ICLUST algorithm was applied to the
368 whole sample. All statistical analyses were performed using the "Psych" package within the R
369 statistical package version 3.1.0 (2014-04-10).

370 *3.3. Comparisons between different short-forms*

371 Short-form and subscale scores were compared to each other, and to SSQ scores. Since the
372 data of the NH participants deviated from normality (Shapiro-Wilk test, $p < 0.001$, kurtosis
373 between 2 and 4 for some variables), these data were compared using non-parametric tests:
374 Wilcoxon's sign test; Friedman's analysis of variance (F-ANOVA); Spearman's correlation
375 coefficient. The data of the HI subjects, and those used for whole-group analyses (including
376 both NH and HI subjects), met the normality assumptions ($|\text{skewness}| < 2.6$; $|\text{kurtosis}| < 2.1$).
377 Accordingly, for these data, parametric tests were used: analysis of variance for repeated
378 measures (R-ANOVA); paired t-tests; Pearson's correlation coefficients. For both NH and HI
379 groups, internal validity of each short-form was assessed using Cronbach's alpha coefficient.
380 Correlations between the SSQ and the different short-forms were calculated. To correct for the
381 hyper-inflation of those correlation coefficients due to the presence of common items in the
382 SSQ and the short-forms, Girard and Christensen's correction for overlapping error variance
383 was used (Girard & Christensen 2008). In order to obtain reciprocal equations allowing to
384 transform the SSQ scores to a short-form score and vice versa, taking into account measurement
385 error of both scores, orthogonal distance regressions were performed on the total population.

386 Multiple-regression analysis of the SSQ, the short-forms, and the subscale scores of the
387 HIN and HIHA subjects were performed using three predictors: Better Ear PTA (in dB HL),
388 Ear Asymmetry in (dB), and Age (in years). The normality of the residuals was checked using
389 the Shapiro-Wilk test, and the assumptions of non-multicollinearity were checked using Dubin-
390 Watson test and VIF statistics. Correlation coefficients were compared using Revelle's two-
391 tailed test for correlated coefficients (R psych package).

392 **3.4. ROC analysis**

393 Receiver-operating-characteristic (ROC) curves were calculated for HI and NH subjects
394 (HI being the patients and NH subjects being the controls), for the SSQ, each of the short-forms,
395 and each of the three subscales, using Robin et al. (2011) R package, pROC. The ROC curves

396 were compared using Venkatraman's test for paired ROCs. The areas under the ROC curve
397 (AUC) and partial AUC across regions of high sensitivity (90 to 100%) were compared using
398 Robin et al.'s bootstrap test, based on the percentile bootstrap method, using 10000 replicates.
399 The Z statistic and two-tailed p value associated with Robin et al.'s bootstrap are given. As the
400 SSQ is designed to assess a difference across two conditions (e.g. HI versus NH, or hearing-aid
401 versus no hearing-aid), the region of high sensitivity appeared to be most relevant.

402

403

RESULTS

1. Factor and cluster analysis of the short-forms

405 To check the underlying factor structure of the various short-forms and their similarity
406 with the SSQ, common factor analysis with maximum likelihood factor extraction was
407 performed on all of the SSQ short-forms except for the SSQ5, as the SSQ5 is for screening
408 purpose and does not reflect the internal three subscales structure of the SSQ. Although the
409 SSQ12 was not meant initially to reflect the three subscales structure of the SSQ, it contains
410 items that were belonging initially to each one of the subscales of the SSQ. Analysis of its
411 internal structure might therefore useful in confirming, or not, that it can reflect the three
412 subscales. This was done separately for the 196 HIN and the 88 HIHA subjects.

1.1. Factor analysis

414 All the results from the factor analyses are in table 1. The Kaiser-Meyer-Olkin (KMO)
415 indexes of Sampling adequacy were above 0.84 (which corresponds to “very good” (Field et al.
416 2012; Fabrigar et al. 1999), with values for individual items all higher than 0.71. Bartlett’s tests
417 of sphericity were all highly significant. The analysis showed systematically three factors with
418 eigenvalues greater than or equal to one (except for the SSQ12s), with cumulative variance
419 explained between 65% (SSQ12s) and 75% (15iSSQ) (71% for the 15iSSQs). The three-factor
420 structure was systematically confirmed by Catell’s scree test, parallel analysis, and Velicer’s
421 MAP criterion. The mean communalities ranged from 0.65 to 0.75. The RMSR ranged between
422 0.03 and 0.04. Hence, a three-factor extraction was always chosen for the final analysis, and an
423 oblique rotation (oblimin) was applied, as the scores of all items were intercorrelated. The three
424 rotated factors explained each at least 21% of the variance (17% for the SSQ12s), and each
425 factor loaded primarily on items corresponding to the main subscales of the SSQ. The three
426 rotated factors correlated with each other, with correlation coefficients ranging from 0.41 to
427 0.67. The different indices showed that this three factor solution was adequate, with a Tucker-

428 Lewis index above 0.9 (one exception at 0.87), a CFI (comparative fit index) above 0.92, and
429 root mean square error of approximation (RMSEA) between 0.08 and 0.13.

430 For the 15iSSQ, the items loadings on each factor showed clear separation between each
431 factor, with no consistent cross-loading across the three factor analysis. A small degree of cross-
432 loading was obtained for #1.5 especially for the 15iSSQs on the 88 HIHA, and for #2.2 but only
433 for 15iSSQ (Fig. 2).

434 For the SSQ15s, the separation between the different loads is less clear: for both
435 samples, item #3.4 showed low communality (0.56) and cross-loading on factor 3 and factor 2
436 and both #3.3 and #2.12 showed a minor degree of cross-loadings on factors 2 and 3. Items #1.5
437 showed some cross-loadings on factors 1 and 2 for the 88 HIHA (Fig. 2). Hence, in the two
438 different populations, items #3.3, #3.4 and #2.12 did not perform as well as the others.

439 For the SSQ12s, the three factors structure is not well respected: the % of variance
440 explained, the communalities were the lowest of all the short forms (table 1). For both samples,
441 #3.14 loads on factor 2, instead on factor 1. Items #3.7, #3.14 and #1.12 showed low
442 communality (<0.53) and heavy cross-loading on factor 1 and factor 2 for #3.14 and #1.12, and
443 on factor 1 and 3 for #3.7. Three groups can be distinguished: a group of 6 items, comprising
444 the #3.14 and the 5 speech items; a group of three items belonging to the spatial scale and a
445 group of three items belonging to the qualities scale (supplemental fig. SF1).

446

447 **1.2. Cluster analysis**

448 Hierarchical cluster analysis (ICLUST) corroborated the three-factor structure, with
449 three distinct clusters corresponding the three subscales (Speech, Spatial and Qualities), for
450 both the 15iSSQ and the SSQ15s forms, but not for the SSQ12s.

451 For the 15iSSQ (Fig. 3), Alpha and beta coefficients were higher than 0.82, which is
452 above their usual criterion values (0.8 for alpha and 0.7 for beta), and the difference between

453 these coefficients was small (<0.1), indicating high homogeneity and consistency within each
454 cluster (Cooksey & Soutar 2006). The goodness-of-fit measure for this three-subscale solution
455 was 0.98, with an RMSR < 0.05 , which corresponds to an excellent fit. The alpha coefficient
456 above 0.91 for the three main sub-clusters showed their high reliability.

457 The spatial items are represented in a close knit cluster (C8), with $\alpha=0.94$, in which
458 subcluster C6 can be identified and represents the items pertaining to localization (locate
459 vehicle, dog, person).

460 Speech items show a close association between items 1.4, 1.11 and 1.6, as a sub-cluster
461 (C5), all pertaining to auditory perception in conditions of several talkers. Items 1.1 and 1.5,
462 concerning talking with one person with one source of noise, are associated in sub-cluster C9.

463 Within the quality items, the two items about voice pitch and familiar music are
464 associated in cluster C3, whereas the other three items, about clarity and naturalness of sounds
465 are associated in sub-cluster C11.

466 For the SSQ15s, the hierarchical cluster analysis showed a three-factor structure, with
467 each factor corresponding to the expected subscale, with an acceptable fit (0.97, RMSR = 0.05),
468 a minimum beta coefficient at 0.77, and a minimum alpha coefficient at 0.81. For the 196 HIN
469 subjects, the sub-cluster “quality” showed a lower reliability ($\alpha= 0.89$, $\beta=0.77$), probably
470 due to the item #3.3 (Music and voice as separate items), whose addition to the cluster led to a
471 decrease of 0.1 in the beta coefficient, with a difference between alpha and beta coefficient of
472 0.12 for this factor, slightly higher than the expected <0.1 (supplemental fig. SF2).

473 For the SSQ12s, the hierarchical cluster analysis showed a three-factor structure, but
474 that did not correspond to the subscales of the SSQ. The three clusters structure showed an
475 acceptable fit (0.94, RMSR = 0.06), a minimum beta coefficient at 0.74, and a minimum alpha
476 coefficient at 0.81. For the 196 HIN subjects, two three-items clusters were formed: C7 and C4
477 and one six items cluster (C9). C9 is composed of a sub-cluster of 4 items (C6) that shows good

478 homogeneity ($\alpha=0.92$, $\beta=0.89$), and two items (#1.12 and #3.14), the addition of which
479 decreases both the beta and alpha coefficients, so that the end cluster, C9, shows a difference
480 between beta and alpha coefficient greater than 15. This reflects a lack of homogeneity of the
481 cluster, with items #1.12 et #3.14 being outliers (supplemental fig. SF3).

482

483 **2. Comparison between SSQ and short-forms:**

484 To assess the usefulness and validity of a new tool, it is necessary to compare it to a
485 “golden standard”, here, the full scale SSQ, and to other similar tools (here, the already existing
486 short forms). Indeed, it is necessary to assess how different and similar the short-forms
487 characteristics (such as missing answer rates, scores for the global scale and the different
488 subscales, internal reliability) are from the full scale.

489 ***2.1. Missing responses***

490 The number of missing responses was analysed per item for the HIHA subjects
491 (supplemental table ST1c). It ranged from 0 to 14.8%, with an average of 3.1% ($SD = 3.5$)
492 across the 49 items of the SSQ. Those percentages were significantly correlated with the
493 percentages obtained in the HIN group ($r = 0.67$, $p < 0.0001$), with the same items (i.e., items
494 2.14 to 2.16 and 3.16) yielding the greatest number of missing responses for these two groups.
495 The 15iSSQ gave an average of 2% ($SD = 1.4$) and 2% ($SD = 2\%$) for the 15iSSQs, with a
496 significant correlation between the two ($r = 0.72$, $p < 0.003$).

497 ***2.2. Global scores (fig. 4)***

498 F-ANOVA showed highly significant differences between the different scores (full
499 version and short-forms) for the NH subjects ($\chi^2 = 58.3$, $df = 7$, $p < 0.00001$) and for the
500 HIHA subjects ($\chi^2 = 68.4$, $df = 7$, $p < 0.00001$) (Fig. 4)

501 For NH subjects, the newly developed short-forms gave significantly greater scores than
502 the full scale, with $W = 163$, $p < 0.02$ for the 15iSSQ, and $W = 113$, $p < 0.0006$ for the 15iSSQs.

503 SSQ12s scores were significantly lower than the SSQ scores ($W=563$, $p<0.0001$). No
504 significant differences were obtained for the SSQ15s or the SSQ5s.

505 For HIHA subjects, all of the short-forms except SSQ5s gave significantly lower scores
506 than the full-form, with $W = 2517$, $p < 0.03$ for the 15iSSQ, $W = 2985$, $p < 0.0001$ for the
507 SSQ15s, and $W = 3698$, $p < 0.00001$ for the SSQ12s.

508 The 15iSSQ and 15iSSQs gave slightly greater cut-off scores than the SSQ in both NH
509 samples (6.1 vs 6.0 for 35 NH), the SSQ15s (5.96) and the SSQ12s (5.31).

510 **2.3. Subscale scores (Fig. 4)**

511 For NH subjects, no statistically significant difference was obtained between the full
512 form and the different short-forms for the Speech ($\text{Chi}^2 = 11.6$, $\text{df} = 5$, $p<0.05$) and Spatial
513 subscales ($\text{Chi}^2 = 6.3$, $\text{df} = 5$, $p > 0.20$). For HIN subjects, no statistically significant difference
514 was obtained for the Spatial subscale (F-ANOVA, $\text{Chi}^2 = 10$, $\text{df} = 5$, $p = 0.07$).

515 However, for the Speech subscale and the HIN subjects, all the different short-forms
516 gave significantly lower scores than the full subscale (F-ANOVA $\text{chi}^2=81.5$, $\text{df}=5$, $p<0.00001$),
517 with $W = 2728$, $p < 0.0006$ for the 15iSSQ, and $W = 3606$, $p < 0.00001$ for the SSQ15s.

518 For the Qualities subscale, highly significant differences between full form and the
519 different short-forms were observed for NH subjects and for HIHA subjects (F-ANOVA, Chi^2
520 $= 68$, $\text{df} = 5$, $p < 0.00001$ for NH and $\text{Chi}^2 = 73$, $\text{df} = 5$, $p < 0.00001$ for HIHA), albeit in
521 opposite directions: for NH subjects, all of the different short-forms yielded significantly
522 greater Qualities-subscale scores than the full form, with $W = 4$, $p < 0.00001$ for the 15iSSQ,
523 and $W = 56$, $p < 0.0001$ for the SSQ15s; for HIHA subjects, the short-forms yielded
524 significantly lower Qualities sub-scores than the full form, with $W = 1071$, $p < 0.0006$ for the
525 15iSSQ, and $W = 741$, $p < 0.00001$ for the SSQ15s.

526 Although the mean scores of the 15iSSQ Speech and Spatial subscales were not lower
527 than the SSQ sub-scores, the cut-off scores of the 15iSSQ Spatial subscale were lower (4.1

528 versus 4.8) than the SSQ Spatial sub-scores, due to larger standard deviations. The SSQ15s
 529 gave the lowest cut-off scores (5.9 and 3.9 for Speech and Spatial respectively). However, the
 530 Qualities cut-off scores were all greater for all short-forms (15iSSQ, SSQ15s) than for the SSQ
 531 (supplemental table ST2).

532 **2.4. Internal validity and correlations between the different SSQ forms.**

533 Cronbach's alpha was higher than 0.83 for SSQ5s, 0.93 for the SSQ12s, and 0.94 for
 534 the SSQ15s and 15iSSQ. Good consistency was also shown within each subscale, with
 535 Cronbach's alphas higher than 0.92 for the Speech and Spatial subscales, and comprised
 536 between 0.85 and 0.91 for the Quality subscale (supplemental table ST3a). Item-to-total
 537 correlations were above 0.67 for the 196 HIN, and above 0.61 for the HIHA (supplemental table
 538 ST3b).

539 The different short-forms correlated very highly with the SSQ, with the lowest
 540 correlation coefficients obtained for the SSQ5s ($r = 0.88$, corrected Pearson) and coefficients
 541 between 0.94 and 0.95 for the other short-forms (Fig. 5). The scores for the three subscales
 542 correlated highly with the corresponding SSQ subscale scores (r above 0.91 for the Speech and
 543 Spatial subscales, and r between 0.78 and 0.85 for the Quality subscale) (supplemental table
 544 ST4a). Similar results were obtained for the non-aided population (supplemental table ST4b).

545 As the relationship between the SSQ and the SSQ12s (and to a lesser degree the SSQ15s,
 546 see Moulin & Richard 2016a) follows a power function rather than a linear one (Noble et al.
 547 2013), we fitted the data for each of the short-forms tested using power functions, according to
 548 Noble's formulae (with $b=1.25$ and $c=0.80$):

$$549 \quad \text{SSQ Short forms} = 10 * (\text{SSQ}/10)^b$$

$$550 \quad \text{SSQ} = 10 * (\text{SSQ Short forms}/10)^c$$

551 Non-linear relationships between scores for the SSQ and the SSQ12s, on the one hand,
 552 and the SSQ and the SSQ15s, on the other hand, were confirmed ($b = 1.215$ [1.180, 1.250] and

553 $b = 1.069 [1.041, 1.097]$, respectively). For the new short-form, the relationship was linear, as
 554 shown by a 95% confidence interval for b overlapping 1 (supplemental table ST5). Similar
 555 analyses performed on the larger sample of hearing-aid non-wearers (NH and HIN) yielded
 556 similar results, with a linear relationship between scores of the SSQ and of the new short-form.

557 Reciprocal equations obtained by orthogonal distance regression between SSQ and the
 558 new short form, using all subjects (i.e., 417) are:

$$559 \quad 15iSSQs = 0.931 * SSQ + 0.492$$

$$560 \quad SSQ = 1.074 * 15iSSQs - 0.529$$

561

562 **3. Construct validity**

563 ***3.1. Influence of hearing-loss***

564 The scores of all the different forms decreased highly significantly with increasing
 565 hearing-loss in the better ear, with a slope ranging from 0.56 (SSQ5s) to 0.81 (15iSSQ) (0.65
 566 for the SSQ) scale points per 10-dB HL. These slopes did not differ significantly from each
 567 other. However, the different short forms differed from each other in the strength of the
 568 correlation between score and HL, with a significantly higher correlation coefficient for the
 569 newly developed short-form than for the full form ($r = 0.51$ versus $r = 0.58$, $t = 2.5$, $p < 0.02$),
 570 and a significantly lower correlation for the SSQ5s than for the full form ($r = 0.41$, $t = 3.6$, $p <$
 571 0.0005).

572 When analysing subscale scores, correlations between scores and HL were not found to
 573 differ significantly across the different SSQ forms (full versus short).

574 Linear regressions performed on the scores of the NH and HIHA subjects showed a
 575 decrease in short-form scores with hearing loss, with a greater slope for the short-forms in
 576 general, except for the SSQ5s.

$$577 \quad SSQ = -0.065 * HL + 8.23 \quad (r = 0.51, 26\% \text{ variance explained})$$

578 $15iSSQ = -0.081*HL+8.47$ ($r = 0.58$, 34% variance explained)

579 All the details of the relationships between scores (for all the different forms and their
580 subscales) and hearing loss are in supplemental table ST6.

581 ***3.2. Influence of patients' characteristics***

582 This analysis sought to investigate predictors of the full- and short-form scores. To this
583 aim, multiple-regression analyses were performed on the scores for each form, using the
584 following explanatory variables: Gender, Age, Better-ear PTA, and (PTA) Asymmetry. The
585 latter two audiometric variables were measured with, and without, hearing-aids, and both
586 measurements were entered as potential predictors in the model. For these analyses, the data of
587 the 88 HIHA subjects were used. Gender and Age were never found to be statistically
588 significant predictors. The best models included hearing thresholds measured with hearing-aids,
589 and PTA Asymmetry measured without hearing-aids. Therefore, multiple-regression results are
590 presented using the better-ear PTA measured with hearing-aids, and PTA asymmetry measured
591 without hearing-aids. The size of the effect (r^2), and the beta coefficients that allow the
592 comparison of the relative influence of each statistically significant predictor on the SSQ scores
593 are summarized in supplemental table ST7. The different short-forms yielded results similar to
594 the full SSQ, with 32% to 39 % of variance explained by the two predictors: Better-ear PTA
595 (beta ranging from -0.28 to -0.35) and PTA Asymmetry (beta ranging from -0.24 to -0.19) for
596 the seven different SSQ forms (SSQ5s, SSQ5ws, SSQ12s, SSQ15s, i15SSQ, i15SSQs, SSQ).
597 Ear asymmetry was not a significant predictor of SSQ12 scores, but was the only significant
598 predictor of SSQ5 scores.

599 The three subscales of the short-forms gave similar results to the three subscales of the
600 SSQ: Better-ear PTA and Asymmetry were significant predictors. For all forms, the greatest
601 dependencies were between the Asymmetry predictor and scores of the Spatial subscale, and
602 between Better-ear PTA and scores of the Quality subscale. The differential scores between the

603 main subscales (Qualities – Spatial; Qualities – Speech) did not show any significant
604 dependency on the two predictors (supplemental table ST7b).

605 Similar analyses were performed on the 196 non-hearing-aid wearers. Results showed
606 that the different short-forms yielded results similar to the SSQ, with 37% to 38% of variance
607 explained by the two predictors: Better Ear PTA (beta from -0.54 to -0.58) and Ear Asymmetry
608 (Beta = -0.29 to -0.32) for the five short-forms. The three subscales of the SSQ15s and 15iSSQ
609 gave results similar to the three subscales of the SSQ, with the Better-ear PTA and PTA
610 Asymmetry as significant predictors. The relative importance of PTA Asymmetry for the
611 Spatial subscale was higher for the 15iSSQs, with an even greater beta coefficient (-0.47) than
612 for hearing loss (-0.44). The differential scores (Qualities – Spatial) had PTA Asymmetry as
613 the main predictor, with 21% of variance explained for the 15iSSQs, which is substantially
614 greater than the same score using the SSQ (11%) or the SSQ15s (10%) (supplemental table
615 ST7b).

616 **3.3. ROC analysis**

617 For the SSQ, the area under the ROC curve (AUC) was equal to 77.3 for the HIN group,
618 and to 80.1 for the HIHA group. For the HIHA group, significantly larger AUCs (Fig. 6) were
619 obtained for the 15iSSQ (AUC=84) and the 15iSSQs (AUC=83) than for the SSQ ($Z = 1.75$, p
620 < 0.08 , and $Z = 2.02$, $p < 0.05$ for the 15iSSQ and 15iSSQs, respectively). By contrast, AUCs
621 for the SSQ12s and the SSQ5s were not significantly larger than the SSQ AUC. For the HIN
622 group, differences in AUC between the SSQ and the different short-forms did not reach
623 statistical significance (supplemental table ST8a, b).

624 Similar analyses were performed for the three subscales (Fig. 6). For the Speech
625 subscale, the different short-forms all showed a larger AUC (> 86.4) than the full-form AUC
626 (84.2), with significant differences for the 15iSSQs ($Z = 2.9$, $p < 0.004$ for the HIHA group; Z
627 $= 2.4$, $p < 0.02$ for the HIN group). For the Spatial subscale, no significant difference between

628 the different short-forms and the full-form were obtained for the HIN group; however, for the
629 HIHA group, the short-forms all produced a smaller AUC (71.9 for the 15iSSQ and 69.3 for
630 the SSQ15s) than the SSQ (72.1), with a significant difference for the 15iSSQs ($Z = -2.2$, $p <$
631 0.03). For the Qualities subscale, the AUC for the 15iSSQs ($Z = -4.4$, $p < 0.0001$) was
632 significantly smaller than the AUC for the full-form for the HIN group, but not for HIHA group.
633 (supplemental tables ST8a and ST8b).

634 AUCs for the spatial subscale were systematically smaller than AUCs for the other
635 subscales, while the AUCs for the Speech subscale were systematically larger than AUCs for
636 the other subscales: this pattern was seen for the SSQ, as well as for the different short-forms.

637 A similar analysis was performed for the high-sensitivity region (sensitivity above
638 90%). For the HIHA group, the newly developed short-forms showed a significantly and
639 systematically larger partial-AUC than the SSQ ($Z = 2.5$, $p < 0.02$ for the 15iSSQ, and $Z = 2.4$,
640 $p < 0.02$ for the 15iSSQs). No significant differences were obtained between the SSQ and the
641 SSQ15s, the SSQ12s, or the SSQ5s. No significant differences were obtained for the HIN group
642 either.

643

644

DISCUSSION

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

The main goal of this study was the creation of a new SSQ short-form using a rigorous data-driven approach, in agreement with the criteria suggested by Stanton et al. (2002) and the factorial-integrity check suggested by Widaman et al. (2011) and Smith et al. (2000). The second goal was a detailed analysis of the SSQ (and its short-forms) using types of analysis not yet reported (for instance, ROC analysis, and internal structure of the short-forms), and involving two different samples (HIN and HIHA patients). By using data from multiple studies, we tried to avoid the caveat of building a short-form specific to a single data set (Widaman et al. 2011). By using various types of criteria for selecting items, we tried to overcome the caveats of a single-criterion approach. For instance, by using only internal-consistency criteria, or by over-using such criteria, one can end up with too narrowly focused a short-form, which explores only some aspects of the full form (Stanton et al. 2002). Following Stanton et al. (2002), we considered three different categories of criteria: “judgmental” (such as the percentage of missing answers, the belonging of the items to another short form, or the readability), “external qualities” (percentage of variance explained by the five main predictors), and “internal qualities” (communalities and factor loadings of factor analysis obtained in two different studies; normative data obtained in NH subjects; inter-item and item-to total correlations). Lastly, we validated the new 15iSSQ using two independent and different samples of participants, including both NH and HIHA, hence avoiding the last two of the nine “sins” of short form building described by Smith et al. (2000): not administering the short and full forms independently, and not using independent samples for validation.

665

1. Construct validity

666

667

668

The internal structure of the short-forms was cross-checked using factor and cluster analyses. The results revealed three clusters for the SSQ15s and the 15iSSQ, each corresponding to a main subscale of the SSQ (Speech, Spatial, and Qualities). However, internal

669 structure of the SSQ12s did not reflect the three subscales of the SSQ, with items belonging to
670 another subscale than expected, and two items showing cross-loading between two factors. This
671 was confirmed by cluster analysis, with one cluster of six items, itself composed of one cluster
672 of four items plus two “outlier items”, and two other clusters of three items. This was expected
673 as the SSQ12 was not built to reflect the three subscales structure of the SSQ, but to reflect the
674 SSQ as a whole (Noble et al. 2013) and, in particular, the 10 pragmatic subscales of the SSQ
675 (Gatehouse and Akeroyd, 2006). The 10 pragmatic subscales group items by meaning, and the
676 SSQ12 items belong to 9 out of 10 of the pragmatic subscales (supplemental table ST1). This
677 is why we did not calculate the scores of the SSQ12s per main subscale, as it was not relevant¹.

678 Factor and hierarchical cluster analyses confirmed that the 15iSSQ has the same three-
679 factor structure as the full SSQ, with no consistent cross-loading. The results further indicate
680 than the new 15iSSQ is superior to the other existing 15-item SSQ short-form, the SSQ15s
681 (Kiessling et al. 2011). In particular, from a construct-validity perspective, factor and
682 hierarchical cluster analyses indicated less cross-loading, and better fits to the three-factor
683 structure of the original SSQ, for the 15iSSQ (and the 15iSSQs) than for the SSQ15s.

684 **2. External validity**

685 Regarding external validity, consistent with the full SSQ, scores for the 15iSSQ
686 decreased with increasing hearing loss, with a higher slope for the 15iSSQ (0.81 points per 10
687 dB HL, 0.75 points per 10 dB HL for the 15iSSQs) than for the full SSQ (0.65 points per 10 dB
688 HL). This outcome can be explained by one of our criteria for item selection: the presence of
689 relatively high scores and lower variability in NH subjects. Although SSQ scores have been
690 initially interpreted as if young NH subjects should have perfect scores (e.g., House et al. 2010),

¹ SSQ12s results per main sub-scales are nevertheless presented in supplemental tables ST2, ST6 and ST7.

691 in reality, actual SSQ scores for NH are often lower than 10 (Demeester et al. 2012; Moulin et
692 al. 2015). Indeed, in the present study, the average score of NH respondents for the Spatial
693 subscale was lower than 8 (see Fig. 4). The 15iSSQ scores were found to be significantly greater
694 than the full-SSQ scores for both NH samples, and significantly smaller for the HIN and HIHA
695 subjects, thus reinforcing the contrast between NH and HI subjects. In addition, the expected
696 correlation between PTA and score was significantly higher for the 15iSSQ (21% of variance
697 explained) than for the SSQ (11% of variance explained). Unfortunately, our data sample is not
698 sufficient to test for more subtle differences in correlations. Indeed, tests of the differences in
699 dependent correlations require large data samples: for instance, a difference in correlation
700 coefficient of 0.05 (0.50 versus 0.55) would require 298 subjects (with a power of 90% and a
701 level of significance of 0.05) (Faul et al. 2009). Nevertheless, ROC analysis confirmed this
702 greater sensitivity of the 15iSSQ (and 15iSSQs) to hearing impairment: for both HIHA and HIN
703 groups, the AUC, a measure of the difference between the scores of the NH and HI groups, was
704 significantly larger for the 15iSSQ (and 15iSSQs) than for the SSQ, whereas AUCs obtained
705 using the other short-forms were not (SSQ5s, SSQ12s and SSQ15s).

706 One of the caveats of reducing the number of items in a questionnaire, is the potential
707 increase of inter-subject variability, the scores being less “smoothed out” by an averaging
708 procedure across a small number of items than across a greater number of items. This has
709 potential detrimental consequences on the disability SSQ cut-off scores. This is not the case
710 here, as for the total score, the cut-off scores obtained with the new short-form (both 15iSSQ
711 and 15iSSQs) were consistently larger than the scores obtained with (1) the SSQ and (2) all the
712 other short-forms.

713 A look at the three main subscale scores gives a more complex picture: the cut-off scores
714 for the Speech and Spatial subscales tend to be lower for all short-forms than for the SSQ. For
715 the qualities subscale, all short-form gave greater cut-off scores than the SSQ qualities. Indeed,

716 the 15iSSQ and 15iSSQs (and to a lesser degree, the SSQ15s) gave larger scores for the
717 Qualities subscale in all samples, especially in NH. For the Spatial and Speech subscales,
718 15iSSQ scores were similar to SSQ scores for both NH samples, but were significantly lower
719 than SSQ scores for HIN and HIHA. As a result, the contrast between Qualities and Speech
720 sub-scores and, to a lesser degree, between Speech and Spatial sub-scores, was higher for HI
721 with the 15iSSQ than with the SSQ. The lower Spatial sub-scores for the HI participants may
722 be attributed to our choice of favoring items that were more sensitive to Ear Asymmetry. This
723 can be observed in the measured dependency of the Spatial scores to Best Ear PTA and Ear
724 Asymmetry, where both predictors contributed almost equally to the scores of the 15iSSQs; by
725 contrast, for the SSQ15s and the SSQ, the contribution of the Best Ear PTA was always
726 substantially greater than the contribution of Ear Asymmetry. The statistical relationship
727 between PTA asymmetry and the differential sub-score, Qualities – Spatial, was substantially
728 stronger for the 15iSSQs (21% of variance explained) than for the SSQ15s (10% of variance
729 explained) or the SSQ (11% of variance explained). Hence, although our selection of items for
730 the new short-form was guided primarily by the imperative to maintain the three main subscales
731 of the SSQ, two advantageous by-products of this selection are a better contrast between NH
732 and HI, and a Spatial subscale that is more sensitive to Ear Asymmetry. This greater sensitivity
733 to ear asymmetry can be highly advantageous in the analysis of self-reported hearing disabilities
734 linked to ear asymmetry (Vannson et al. 2015), unilateral hearing loss (Olsen et al. 2012; Dwyer
735 et al. 2014; Douglas et al. 2007), and the evaluation of the benefits of hearing rehabilitation
736 strategies of those asymmetrical losses (Pai et al. 2012; Dumper et al. 2009).

737 **3. SSQ/short form relationships and differences**

738 It is worth noting that we found a linear relationship between scores obtained with the
739 new short-form (for both 15iSSQ and 15iSSQs) and scores for the full SSQ. This differs from
740 the SSQ12s for which the relationship is clearly nonlinear (Moulin & Richard 2016b; Noble et

741 al. 2013) and, to a lesser degree, from the SSQ15s, and makes it straightforward to infer scores
742 for the full SSQ based on the new short-form scores, or *vice versa*.

743 One limitation of this study stems from the fact that test-retest reliability of the new
744 short-form was not verified on a large data sample. To the best of our knowledge, this
745 shortcoming applies also to the other SSQ short-forms. However, it appears highly unlikely that
746 the 15iSSQ is any less reliable than its existing 15- and 12-item counterparts, the SSQ15s and
747 the SSQ12s. First, the items in this new short-form come from the same set of items that was
748 used to create the SSQ15s and the SSQ12s, namely the complete set of SSQ questions, which
749 has shown a high degree of correlation between several administration modes and at different
750 time points (Singh & Pichora-Fuller 2010). Second, we obtained very good correlations
751 between the 15iSSQs and the 15iSSQ ($r=0.93$) and no statistically significant differences
752 between 15iSSQs and 15iSSQ in the same samples (NH and HIHA), whether the full score or
753 scores per subscales were considered. Although the delay between the two administrations was
754 too low (less than a day for about 40% of them and less than a week for most of them) to really
755 assess long-term reproducibility, and responses to the subset version (15iSSQs) were, to some
756 degree, influenced by the other questions of the full SSQ, it does give a clear indication of the
757 good test/retest reproducibility of the 15iSSQ.

758 Even if it were merely equivalent to the full SSQ for differentiating between NH and HI
759 subjects, the 15iSSQ could be advantageously used for this purpose in research studies or
760 clinical work, given its shorter length. In fact, the results showed that, in some respects, the
761 short-form affords *better* discrimination between NH and HI subjects than the full SSQ, and a
762 better sensitivity to ear asymmetry thanks to the Spatial subscale. One qualification to this
763 conclusion stems from the finding of a statistically significant smaller AUCs for the 15iSSQ
764 than for the SSQ, for one specific combinations of HI subgroups and subscales: the Qualities
765 subscale with the HIN group. For this reason, we cannot recommend using the 15iSSQ in lieu

766 of the full SSQ in all situations, especially those in which longer length of the latter is not a
 767 major limitation.

768

769 **4. Overall Differences between short-forms**

770 The differences obtained between the different short-forms appear small overall,
 771 especially the differences between the 15iSSQs and the SSQ15s. However, we need to take into
 772 account the fact that all the short-forms have been compared as subsets of a common SSQ
 773 questionnaire, hence a great proportion of data are identical across several short-forms. In
 774 particular, the 15iSSQs and the SSQ15s share 7 items out of 15, i.e., almost 50% of the data are
 775 identical between the two. Nevertheless, importantly, the 15iSSQs outperforms the SSQ15s in
 776 having greater cutoff values for both NH samples, for both the main scores and the three main
 777 subscales, as well as a greater dependency of its spatial subscale on ear asymmetry, a greater
 778 AUC for main scores and speech subscales, and a more defined internal structure. In addition,
 779 the independently applied 15iSSQ exhibits better characteristics than the 15iSSQs.

780 Although the sample sizes in this study and the redundancy of items across different
 781 short-forms, contributed to limit the magnitude of any statistical differences between the
 782 15iSSQ and the other short-forms, we found that the 15iSSQ outperformed the other short-
 783 forms in three main aspects:

- 784 • *An internal structure in three clearly defined subscales, that are the same as the*
 785 *three subscales of the SSQ.* This is not the case with the SSQ12s. The SSQ15s 's
 786 three subscales are less clearly defined, with some cross-loading on two factors in
 787 the factorial analysis.
- 788 • *A significantly greater dependency of the 15iSSQ on hearing impairment, such as*
 789 *hearing loss and ear asymmetry.* This is shown by a significantly stronger
 790 correlation between 15iSSQ scores and hearing loss, the greater percentage of

791 variance explained by hearing-loss and ear asymmetry (for the spatial subscale) than
792 the other short forms, a significantly greater area under the curve for the ROC
793 analysis. These results demonstrate a greater contrast between normal hearing
794 subjects and hearing impaired subjects.

795 • *A linear relationship between SSQ and 15iSSQ*, shown for the 15iSSQs as a subset
796 of the SSQ, and for an independent administration of the 15iSSQ. Such a linear
797 relationship is obtained with the SSQ5s as well, but both the SSQ12s/SSQ and
798 SSQ15s/SSQ relationships are significantly non-linear (power function). A linear
799 relationship allows for easier interchangeability between SSQ and short-form
800 scores.

801

802 **Acknowledgements**

803 The authors wish to thank Jeremy Montagnat-Misson for technical help during the data
804 collection.

REFERENCES

- 805
- 806 Akeroyd, M.A., Guy, F.H., Harrison, D.L., et al. (2014). A factor analysis of the SSQ (Speech,
807 Spatial, and Qualities of Hearing Scale). *Int. J. Audiol.*, 53, 101–114.
- 808 Banh, J., Singh, G., Pichora-Fuller, M.K. (2012). Age affects responses on the Speech, Spatial,
809 and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss. *J. Am.*
810 *Acad. Audiol.*, 23, 81-91; quiz 139-140.
- 811 Cooksey, R.W., Soutar, G.N. (2006). Coefficient Beta and Hierarchical Item Clustering An
812 Analytical Procedure for Establishing and Displaying the Dimensionality and
813 Homogeneity of Summated Scales. *Organ. Res. Methods*, 9, 78–98.
- 814 Demeester, K., Topsakal, V., Hendrickx, J.-J., et al. (2012). Hearing disability measured by the
815 speech, spatial, and qualities of hearing scale in clinically normal-hearing and hearing-
816 impaired middle-aged persons, and disability screening by means of a reduced SSQ (the
817 SSQ5). *Ear Hear.*, 33, 615–616.
- 818 Douglas, S.A., Yeung, P., Daudia, A., et al. (2007). Spatial hearing disability after acoustic
819 neuroma removal. *The Laryngoscope*, 117, 1648–1651.
- 820 Dumper, J., Hodgetts, B., Liu, R., et al. (2009). Indications for bone-anchored hearing AIDS: a
821 functional outcomes study. *J. Otolaryngol. - Head Neck Surg. J. Oto-Rhino-Laryngol.*
822 *Chir. Cervico-Faciale*, 38, 96–105.
- 823 Dwyer, N.Y., Firszt, J.B., Reeder, R.M. (2014). Effects of unilateral input and mode of hearing
824 in the better ear: self-reported performance using the speech, spatial and qualities of
825 hearing scale. *Ear Hear.*, 35, 126–136.
- 826 Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., et al. (1999). Evaluating the use of
827 exploratory factor analysis in psychological research. *Psychol. Methods*, 4, 272–299.

- 828 Faul, F., Erdfelder, E., Buchner, A., et al. (2009). Statistical power analyses using G*Power
829 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods*, 41, 1149–
830 1160.
- 831 Field, A., Miles, J., Field, Z. (2012). *Discovering Statistics Using R* 1 edition., London ;
832 Thousand Oaks, Calif: SAGE Publications Ltd.
- 833 Gatehouse, S., Akeroyd, M. (2006). Two-eared listening in dynamic situations: Audición con
834 dos oídos en situaciones dinámicas. *Int. J. Audiol.*, 45, 120–124.
- 835 Girard, T.A., Christensen, B.K. (2008). Clarifying problems and offering solutions for
836 correlated error when assessing the validity of selected-subtest short forms. *Psychol.*
837 *Assess.*, 20, 76–80.
- 838 Gonzalez, E.C. de M., Almeida, K. de (2015). Cross-cultural adaptation of the Speech, Spatial
839 and Qualities of Hearing Scale (SSQ) to Brazilian Portuguese (Adaptação cultural do
840 questionário Speech, Spatial and Qualities of Hearing Scale (SSQ) para o Português
841 Brasileiro). *Audiol. - Commun. Res.*, 20, 215–224.
- 842 House, J.W., Kutz, J.W., Chung, J., et al. (2010). Bone-anchored hearing aid subjective benefit
843 for unilateral deafness. *The Laryngoscope*, 120, 601–607.
- 844 John, O.P., Soto, C.J. (2009) The Importance of Being Valid: Reliability and the Process of
845 Construct Validation. In *Handbook of Research Methods in Personality Psychology*,
846 Robins, R.W., Fraley, R.C., Krueger, R.F. eds. 1 edition., New York, NY London: The
847 Guilford Press.
- 848 Kiessling, J., Grugel, L., Meister, I.G., et al. (2011). Übertragung der Fragebögen SADL,
849 ECHO und SSQ ins Deutsche und deren Evaluation. German translations of
850 questionnaires SADL, ECHO and SSQ and their evaluation. *Z Audiol*, 6–16.

- 851 Kim, B.J., An, Y.-H., Choi, J.-W., et al. (2017). Standardization for a Korean Version of the
852 Speech, Spatial and Qualities of Hearing Scale: Study of Validity and Reliability.
853 *Korean J. Otorhinolaryngol.-Head Neck Surg.*, 60, 279–294.
- 854 Loewenthal, K.M. (2001). *An Introduction to Psychological Tests and Scales*, Psychology
855 Press.
- 856 Moulin, A., Pauzie, A., Richard, C. (2015). Validation of a French translation of the speech,
857 spatial, and qualities of hearing scale (SSQ) and comparison with other language
858 versions. *Int. J. Audiol.*, 54, 889–898.
- 859 Moulin, A., Richard, C. (2016a). Sources of variability of speech, spatial, and qualities of
860 hearing scale (SSQ) scores in normal-hearing and hearing-impaired populations. *Int. J.*
861 *Audiol.*, 55, 101–109.
- 862 Moulin, A., Richard, C. (2016b). Validation of a French-Language Version of the Spatial
863 Hearing Questionnaire, Cluster Analysis and Comparison with the Speech, Spatial, and
864 Qualities of Hearing Scale. *Ear Hear.*, 37, 412–423.
- 865 Noble, W., Jensen, N.S., Naylor, G., et al. (2013). A short form of the Speech, Spatial and
866 Qualities of Hearing scale suitable for clinical use: The SSQ12. *Int. J. Audiol.*, 52, 409–
867 412.
- 868 Olsen, S.Ø., Hernvig, L.H., Nielsen, L.H. (2012). Self-reported hearing performance among
869 subjects with unilateral sensorineural hearing loss. *Audiol. Med.*, 10, 83–92.
- 870 Pai, I., Kelleher, C., Nunn, T., et al. (2012). Outcome of bone-anchored hearing aids for single-
871 sided deafness: A prospective study. *Acta Otolaryngol. (Stockh.)*, 132, 751–755.
- 872 Putnam, S.P., Rothbart, M.K. (2006). Development of Short and Very Short Forms of the
873 Children’s Behavior Questionnaire. *J. Pers. Assess.*, 87, 102–112.
- 874 Revelle, W. (1979). Hierarchical Cluster Analysis And The Internal Structure Of Tests.
875 *Multivar. Behav. Res.*, 14, 57–74.

- 876 Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale
877 construction. *Behav. Res. Methods Instrum.*, 10, 739–742.
- 878 Robin, X., Turck, N., Hainard, A., et al. (2011). pROC: an open-source package for R and S+
879 to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.
- 880 Singh, G., Kathleen Pichora-Fuller, M. (2010). Older adults' performance on the speech,
881 spatial, and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of
882 interview and self-administration methods. *Int. J. Audiol.*, 49, 733–740.
- 883 Smith, G.T., McCarthy, D.M., Anderson, K.G. (2000). On the sins of short-form development.
884 *Psychol. Assess.*, 12, 102–111.
- 885 Stanton, J.M., Sinar, E.F., Balzer, W.K., et al. (2002). Issues and Strategies for Reducing the
886 Length of Self-Report Scales. *Pers. Psychol.*, 55, 167–194.
- 887 Vannson, N., James, C., Fraysse, B., et al. (2015). Quality of Life and Auditory Performance
888 in Adults with Asymmetric Hearing Loss. *Audiol. Neurotol.*, 20, 38–43.
- 889 Widaman, K.F., Little, T.D., Preacher, K.J., et al. (2011). On creating and using short forms of
890 scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas,
891 eds. *Secondary data analysis: An introduction for psychologists*. (pp. 39–61).
892 Washington, DC, US: American Psychological Association.
- 893 Zinbarg, R.E., Revelle, W., Yovel, I., et al. (2005). Cronbach's α , Revelle's β , and McDonald's
894 ω^2 : their relations with each other and two alternative conceptualizations of reliability.
895 *Psychometrika*, 70, 123–133.

FIGURES

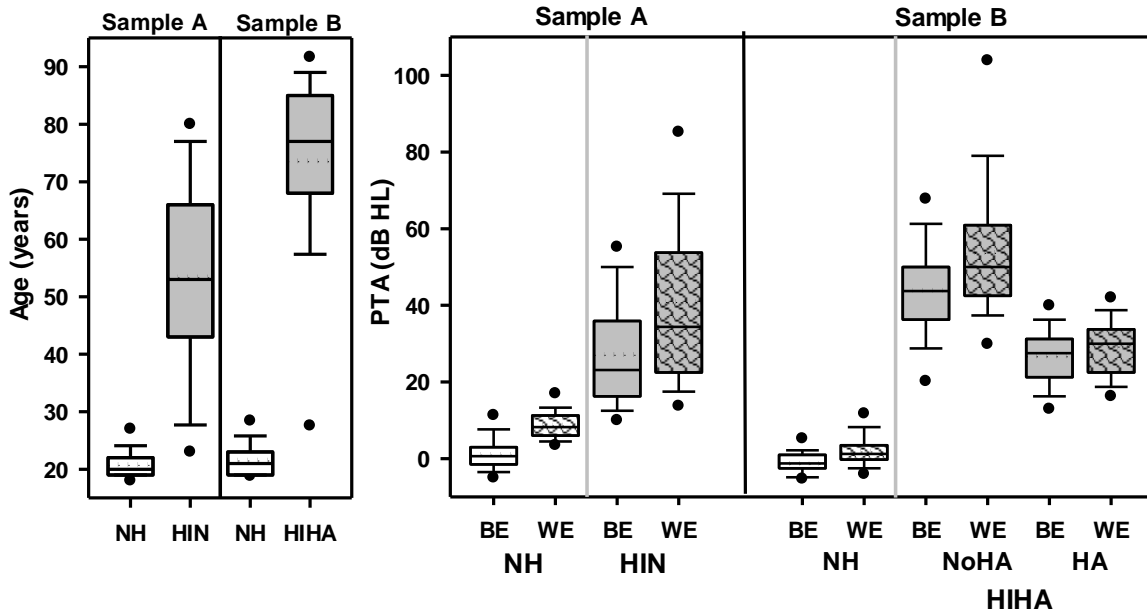


Figure 1

Statistical characteristics of the different subject samples. Left panel: age. Middle and right panels: best-ear (BE) and worst-ear (WE) PTA. For Sample B, all HI subjects were hearing-aid wearers, and PTAs measured with HAs and without HAs ('No HA') are reported.

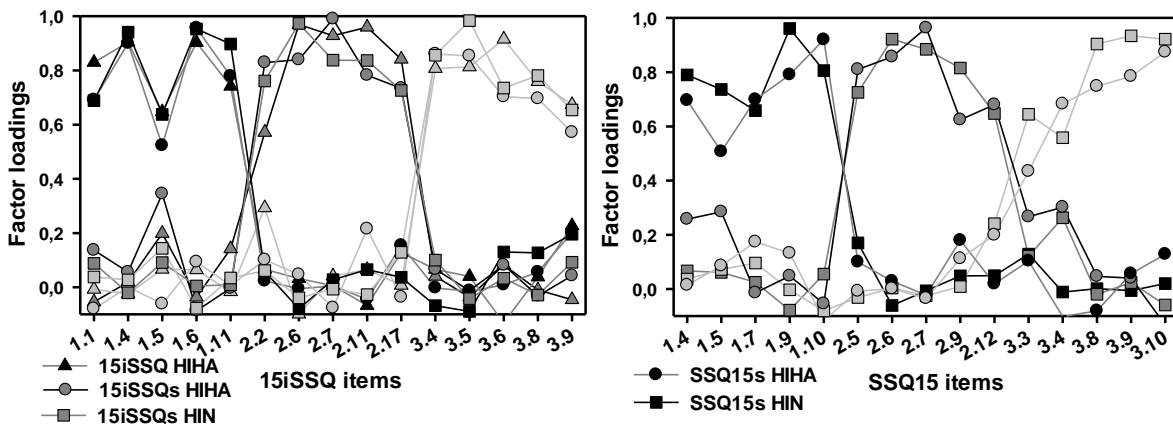


Figure 2

Factor loadings computed using factor analysis applied on the responses to 15-item subsets from the full SSQ corresponding to the 15iSSQ (left panel) or to the SSQ15s (right panel). The three factors are indicated by different colors (black: factor 1; dark-grey: factor 2; light-grey: factor 3). The different subject samples are shown using different symbols, as indicated on the figure: HIHA: hearing impaired hearing aid (n=88) and HIN: hearing impaired without hearing aids (n=196). '15iSSQs' refers to data collected separately in 88 HIHA subjects using the 15iSSQ.

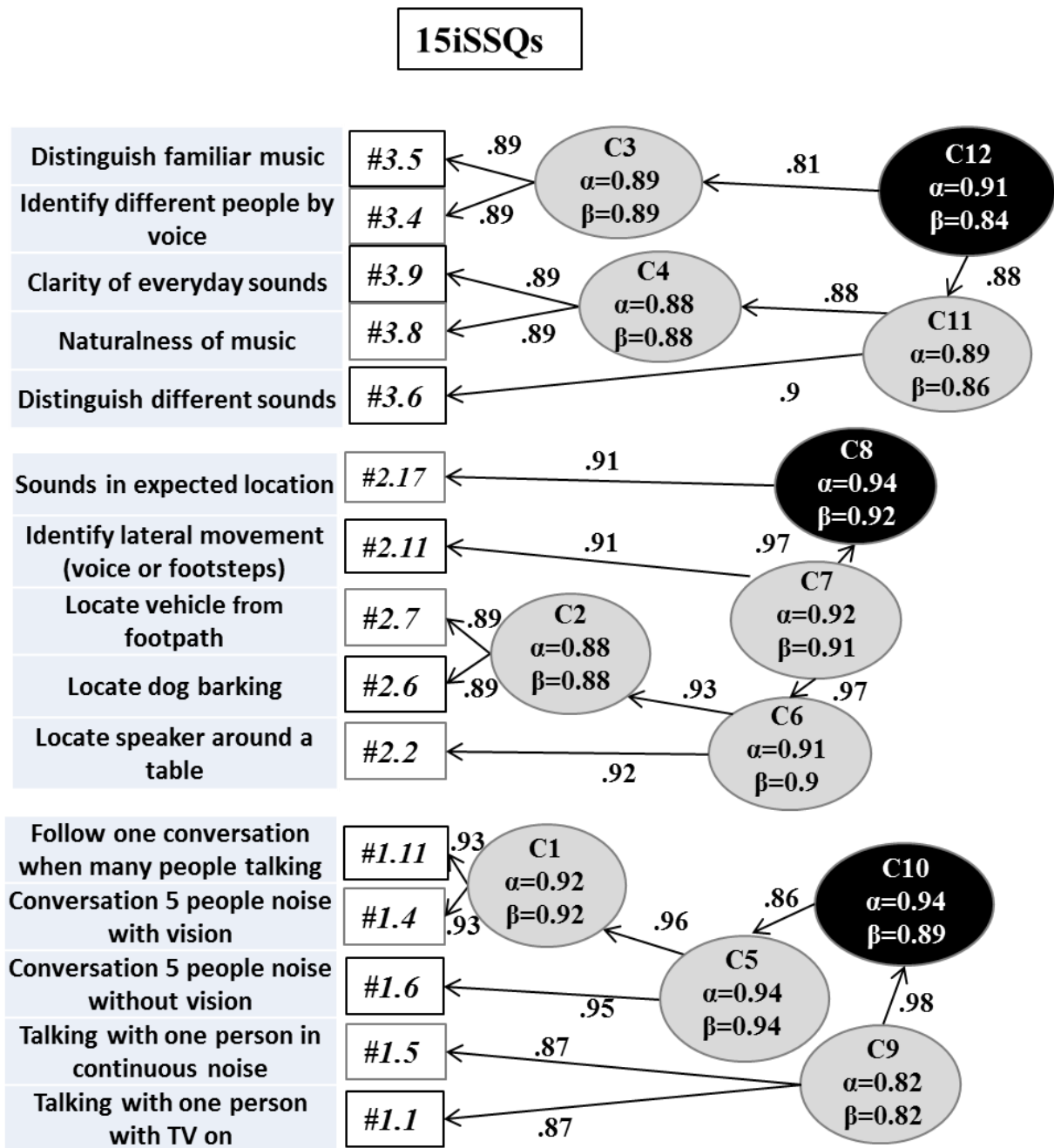


Figure 3

Tree diagram obtained using hierarchical cluster analysis (ICLUST) of the 15iSSQs data from 196 HI subjects. Shortened verbal descriptions (after Bahn et al., 2012) of the 15 items are listed on the left, grouped by subscale. The most similar items are combined first, and increasingly less similar clusters are represented from left to right. For each cluster, the alpha coefficient and Revelle’s B (worst-split-half reliability) are provided. Three main clusters (in black: C8, C10 and C12), corresponding to a main subscale, are identified.

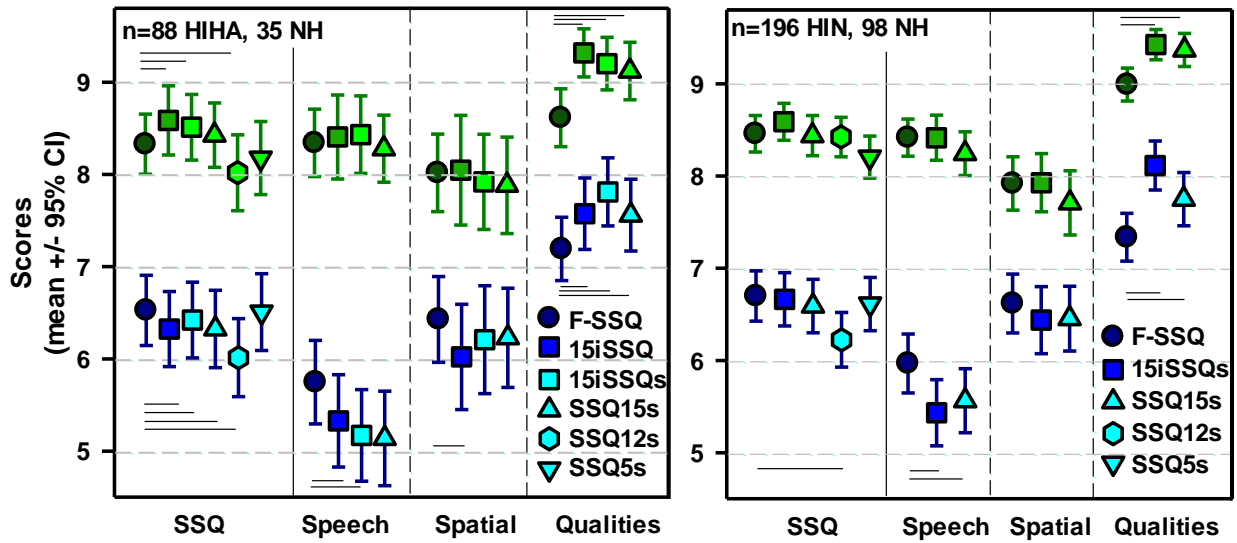


Figure 4

Left panel: full and short-form SSQ scores for the 88 HIHA subjects (blue) and 35 young NH subjects (green) composing Sample B. The short-form scores were either, obtained using the 15iSSQ or, computed by tallying scores across five, 12, or 15-item subsets of the (full) SSQ corresponding to the SSQ5s, the SSQ12s, the SSQ15s, or the 15iSSQ; the latter sub-scores are referred to as ‘15iSSQs’, to distinguish them from scores obtained using the 15iSSQ. Right panel: full and short-form SSQ scores obtained using subsets of the (full) SSQ in the 196 HI subjects (non hearing-aid wearers: HIN, blue) and 98 young NH subjects (green) composing Sample A. Error bars show 95% confidence intervals. Statistically significant differences between SSQ and the different short-forms as represented as horizontal bars (Wilcoxon tests, $p < 0.05$, see text for details).

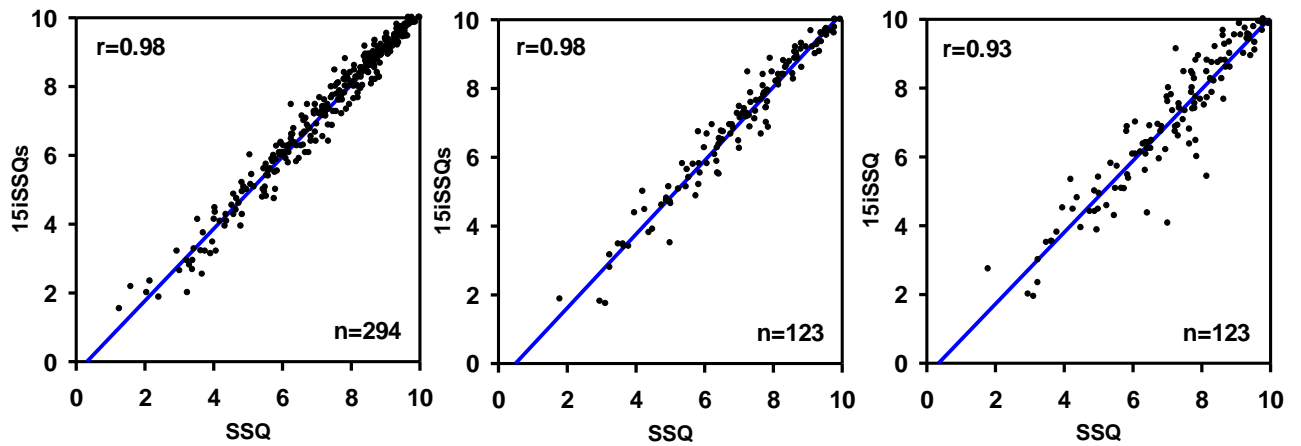


Figure 5

Scatterplots of the scores obtained for the 15iSSQ as a function of the full SSQ score, for the 15iSSQ taken as (1) a subset of the SSQ in sample A data (left panel); (2) a subset of the SSQ in sample B data (middle panel); (3) an independent administration of the 15iSSQ in the same population as in (2) (right panel). Linear regressions are represented by blue lines and Pearson's correlation coefficient (r) is reported for each figure.

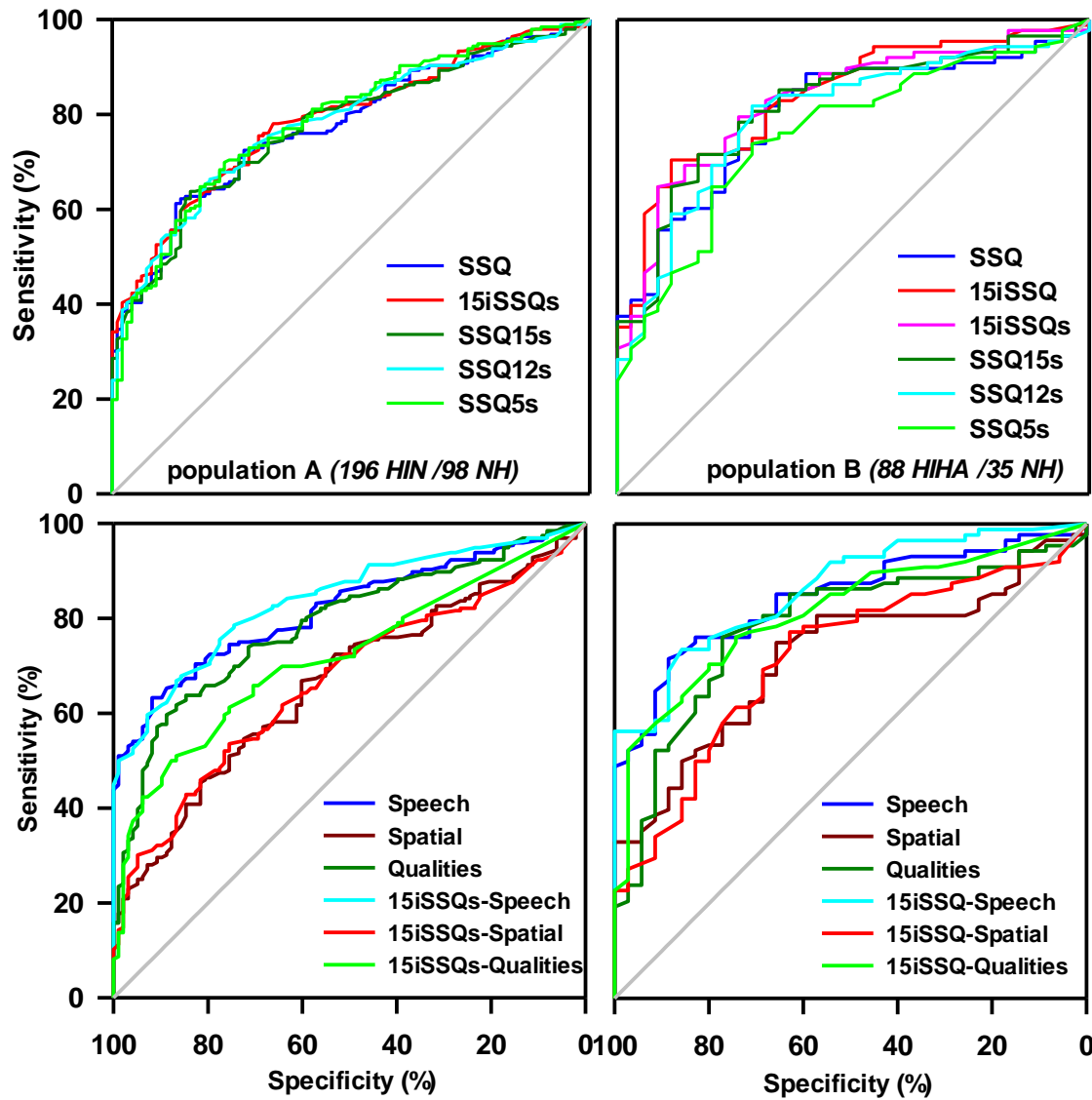


Figure 6

ROCs for each subscale (Speech, Spatial, Qualities). Left: ROCs computed based on Sample-A data. Right: ROCs computed based on Sample-B data. Top row: ROCs computed on scores for the (full) SSQ and for various SSQ short-forms. Bottom row: ROC curves computed on subscale scores for either the (full) SSQ or the 15iSSQ.

TABLE

Population analysed	88 HIHA patients				196 HI patients		
	15iSSQ	15iSSQs	SSQ15s	SSQ12s	15iSSQs	SSQ15s	SSQ12s
Kaiser-Meyer-Olkin (kmo) index	0.92	0.84	0.9	0.88	0.93	0.9	0.91
Minimum KMO value for individual items	0.86	0.71	0.73	0.79	0.88	0.86	0.88
Bartell's test of sphericity (df=105, p<0.00001) (Chi2)	1300	1220	1069	676	2838	2656	1729
Eigenvalues of first 3 factors (F1, F2 and F3)	8.2, 1.7, 1.4	8.2, 1.3, 1.1	8.0, 1.3, 1.0	4.0, 3.3, 0.6	8.8, 1.5, 1.0	8.6, 1.2, 1.0	6.2, 1.2, 0.8
% variance explained by the 3 factors	55, 11, 9	55, 9, 8	54, 9, 7	33,27,5	58, 10, 7	58, 8, 7	52, 10, 6
Cumulative variance explained (%)	75	71	69	65	75	72	68
Mean communality (SD)	0.75 (0.12)	0.71, (0.15)	0.69, (0.10)	0.65 (0.18)	0.75 (0.08)	0.71 (0.10)	0.68 (0.16)
Minimum communality	0.59	0.49	0.55	0.42	0.6	0.55	0.41
% variance explained by the 3 rotated factors	28, 24, 24	27, 23, 21	27, 21, 21	29,20,17	26, 25, 24	25, 24, 23	30, 20, 18
Correlation coefficient between the 3 factors	0.44 to 0.63	0.52 to 0.67	0.45 to 0.59	0.41 to 0.64	0.54 to 0.66	0.65 to 0.66	0.62 to 0.67
Items on which F1 loads, with minimum load	Spatial, 0.57	Spatial, 0.73	Spatial, 0.62	Qualities, 0.32	Speech, 0.64	Spatial, 0.65	Qualities, 0.28
Items on which F2 loads, with minimum load	Speech, 0.65	Speech, 0.52	Speech, 0.51	Speech, 0.47	Quality, 0.65	Quality, 0.56	Speech, 0.39
Items on which F3 loads, with minimum load	Quality, 0.67	Quality, 0.57	Quality, 0.43	Spatial, 0.55	Spatial, 0.73	Speech, 0.66	Spatial, 0.73
Tucker-Lewis index	0.96	0.87	0.9	0.95	0.93	0.92	0.95
CFI (Comparative fit index)	0.98	0.92	0.94	0.98	0.96	0.95	0.97
RMSR	0.03	0.04	0.04	0.04	0.03	0.03	0.03
RMSEA (root mean square error of approximation)	0.08	0.13	0.11	0.08	0.1	0.1	0.08

Table 1

Results of factor analyses performed on 15iSSQ, 15SSQs, SSQ15s, SSQ12s data from the HIHA and HIN subjects.

List of SDC:

Appendix : **SupplementaryData.xlsx** which contains all the supplementary tables (ST1 to ST8) and the supplementary figures (SF1 to SF3).