



HAL
open science

Auditory and Audiovisual Close-shadowing in Post-Lingually Deaf Cochlear-Implanted Patients and Normal-Hearing Elderly Adults

Lucie Scarbel, Denis Beaudemps, Jean-Luc Schwartz, Marc Sato

► **To cite this version:**

Lucie Scarbel, Denis Beaudemps, Jean-Luc Schwartz, Marc Sato. Auditory and Audiovisual Close-shadowing in Post-Lingually Deaf Cochlear-Implanted Patients and Normal-Hearing Elderly Adults. *Ear and Hearing*, 2018, 39 (1), pp.139-149. 10.1097/AUD.0000000000000474 . hal-01546756v2

HAL Id: hal-01546756

<https://hal.science/hal-01546756v2>

Submitted on 2 Jul 2017

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.

Auditory and Audiovisual Close-shadowing in Post-Lingually Deaf Cochlear-Implanted Patients and Normal-Hearing Elderly Adults

Lucie Scarbel¹, Denis Beautemps¹, Jean-Luc Schwartz¹ & Marc Sato²

¹ GIPSA-LAB, Département Parole & Cognition, CNRS & Grenoble Université, Grenoble, France

² Laboratoire Parole & Langage, CNRS & Aix-Marseille Université, Aix-en-Provence, France

Financial Disclosures/Conflict of Interest

This research was funded by the French National Agency for Research (ANR) and a European research council advanced grant (ERC)
The authors declare no competing financial interests.

Address correspondence to Lucie Scarbel, GIPSA-LAB, UMR CNRS 5216, Grenoble Université, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hères, France. Tel: (+33) (0)476 824 128. Fax: (+33) (0)476 574 710. E-mail: lucie.scarbel@gipsa-lab.grenoble-inp.fr.

1 **ABSTRACT**

2 Objectives: The goal of this study was to determine the impact of auditory deprivation
3 and age-related speech decline on perceptuo-motor abilities during speech processing in
4 post-lingually deaf cochlear-implanted participants and in normal-hearing elderly
5 participants.

6 Design: A close-shadowing experiment was carried out on ten cochlear-implanted
7 patients and ten normal-hearing elderly participants, with two groups of normal-hearing
8 young participants as controls. To this end, participants had to categorize auditory and
9 audiovisual syllables as quickly as possible, either manually or orally. Reaction times and
10 percentages of correct responses were compared depending on response modes, stimulus
11 modalities and syllables.

12 Results: Responses of cochlear-implanted subjects were globally slower and less accurate
13 than those of both young and elderly normal-hearing people. Adding the visual modality was
14 found to enhance performance for cochlear-implanted patients, whereas no significant
15 effect was obtained for the normal-hearing elderly group. Critically, oral responses were
16 faster than manual ones for all groups. In addition, for normal-hearing elderly participants,
17 manual responses were more accurate than oral responses, as was the case for normal-
18 hearing young participants when presented with noisy speech stimuli.

19 Conclusions: Faster reaction times were observed for oral than for manual responses in
20 all groups, suggesting that perceptuo-motor relationships were somewhat successfully
21 functional after cochlear implantation, and remain efficient in the normal-hearing elderly
22 group. These results are in agreement with recent perceptuo-motor theories of speech
23 perception. They are also supported by the theoretical assumption that implicit motor

24 knowledge and motor representations partly constrain auditory speech processing. In this
25 framework, oral responses would have been generated at an earlier stage of a sensorimotor
26 loop, whereas manual responses would appear late, leading to slower but more accurate
27 responses. The difference between oral and manual responses suggests that the perceptuo-
28 motor loop is still effective for normal-hearing elderly subjects, and also for cochlear-
29 implanted participants despite degraded global performance.

30

31 **KEY-WORDS**

32 Speech perception, speech production, close-shadowing, cochlear implant, deafness,
33 elderly, sensory-motor interactions.

34

35

36

INTRODUCTION**37 Close-shadowing as a paradigm for the study of perceptuo-motor interactions**

38 Speech communication can be viewed as an interactive process involving a functional
39 coupling not only between the listener's sensory system and the speaker's motor system,
40 but also inside the brain of each interlocutor. The neuroanatomical architecture of sensory-
41 motor coupling in the human brain has been conceptualized in the so-called dual stream
42 model of speech processing (Hickok & Poeppel, 2007). In this model, a bilateral ventral
43 stream within the temporal lobe, associate auditory areas in the superior temporal regions
44 to lexical and semantic information in the middle/inferior temporal gyrus. In parallel, a
45 dorsal stream, connecting temporal, parietal and frontal areas mainly in the left hemisphere
46 of the brain, is in charge of sensory-motor integration.

47 The functional role of sensory-motor coupling has been explored both in speech
48 production and in speech perception. Concerning speech production, it is now widely
49 accepted that auditory areas provide information towards frontal areas in charge of defining
50 and implementing motor control strategies. Auditory areas would be used both to provide
51 targets to control learning, and to constantly send information online, ensuring robustness
52 to perturbations through auditory feedback (Perkell et al., 2000; Guenther, 2006).

53 Concerning speech perception, conceptions are more subject to debate, even though a
54 consensus is progressively emerging. According to motor theories, such as the motor theory
55 of speech perception (Liberman et al., 1985), or the direct realist theory (Fowler, 1986),
56 sensory-motor linkage would set the basis for motor or direct perception. This would be
57 through the recovery of the speaker's motor intentions or articulatory gestures.
58 Alternatively, auditory theories (e.g. Diehl et al., 2004; Stevens and Blumstein 1978, 1979;

59 Lindblom et al. 1988, 1990) claim that speech perceptual processing and categorization are
60 based on acoustic features and auditory representations. Neurocognitive data acquired over
61 the last ten years have clarified the debate somewhat, by showing that the dorsal route is
62 active in speech perception and that the motor system is involved in speech perception, at
63 least in adverse conditions (see a recent review in Skipper et al., 2017). This has led to the
64 emergence of perceptuo-motor theories of speech perception. These theories propose that
65 both motor and auditory representations and processes would be combined in the human
66 brain to elaborate phonetic decisions (e.g. Skipper et al. 2007; Schwartz et al. 2012).

67 Close-shadowing provides a behavioral paradigm well-suited to assess sensory-motor
68 coupling. In this task, subjects have to repeat a speech stimulus immediately after hearing it.
69 Porter and Castellanos (1980) and Fowler et al. (2003) observed very fast reaction times
70 (RTs) when participants had to shadow a syllable as quickly as possible. Importantly, oral
71 responses were faster than manual ones (Galantucci et al. 2006). This difference led to the
72 theoretical assumption that speech perception involves articulatory gestures, and that
73 gesture perception directly controls speech response and makes it faster. This hypothesis
74 was supported by evidence for convergence effects in shadowing experiments (e.g.
75 Goldinger, 1998; Nye & Fowler, 2003). Visual modality also plays a role, both in speeding
76 responses, particularly for noisy acoustic stimuli (Scarbel et al., 2014), and in increasing
77 convergence effects (Dias & Rosenblum, 2016).

78 Close-shadowing therefore appears to provide “a continuous reflection of the outcome of
79 the process of language comprehension, (and) provides us with uniquely privileged access to
80 the properties of the system” (Marslen-Wilson, 1985, p. 55). However, the interpretation
81 relating the high speed of close-shadowing responses to the articulatory nature of speech
82 percepts e.g. Galantucci et al. (2006) remains controversial. Firstly, it can be suggested that

83 the rapidity of oral responses only reflects a sensory-motor coupling process in which
84 auditory speech decoding would automatically evoke preparatory motor responses. Oral
85 responses would be particularly fast if the task engaged the subjects to answer rapidly, but
86 auditory processing would not necessarily use motor representation in its functioning. This is
87 compatible with the way exemplary models are applied to speech (e.g. Pierrehumbert, 2001)
88 and to how “echos” can be introduced as close-shadowing models in such frameworks
89 (Goldinger, 1998). A second line of controversy is supplied by the observation that close-
90 shadowing involves a certain amount of linguistic processing [related to phonology, syntax
91 and semantics (Marslen-Wilson, 1985; Mitterer & Müsseler, 2013)] leading Mitterer &
92 Müsseler (2013) to conclude that there would be “evidence for a loose perception–action
93 coupling in speech” (p. 557).

94 In this context, a recent close-shadowing study carried out by our group sheds some
95 interesting light on the question. In this study on normal-hearing young participants (NHY),
96 both oral and manual responses were evaluated (Scarbel et al. 2014). Conditions were such
97 that the perception of auditory and audiovisual speech stimuli was assessed, embedded or
98 not in white noise. Whatever the modality of presentation, and in agreement with previous
99 studies, oral responses were much faster than manual ones. However, two new findings
100 were discovered in the presence of acoustic noise. Firstly, as mentioned above, the
101 audiovisual modality led to both faster and more accurate responses than the auditory
102 modality. Secondly, and more surprisingly, it appeared that the oral responses were faster
103 but also less accurate than manual responses.

104 These results are important because they are not compatible either with the hypothesis
105 of a pure gestural or motor encoding of speech perception, or with a simple automatic
106 sensory-motor coupling process. Indeed, if sensory-motor coupling were automatic, for

107 translating sounds into gestures in the first case, or just to produce a rapid oral response in
108 the second one, there would be no difference in response accuracy between oral and
109 manual responses.

110 A possible interpretation in the framework of perceptuo-motor theories of speech
111 perception may be provided in the context of analysis-by-synthesis models. In the
112 feedforward-feedback neurobiological model of speech perception of Skipper et al. (2007),
113 articulatory motor representations would be internally generated to partly constrain
114 phonetic interpretation of the sensory input through the internal generation of candidate
115 articulatory categorizations. We suggested that oral replies might be generated at an early
116 stage in the perceptuo-motor loop, therefore providing faster, but possibly less accurate,
117 answers, particularly in the presence of noise. Another interpretation in the context of
118 automatic generation of oral responses would be that such feedback would be generated
119 early in the auditory perception process. It would thus provide an early window on
120 perceptual processing, at a stage where a final decision would still be inaccurate.

121 Whatever the interpretation, the data of Scarbel et al. (2014) show that close-shadowing
122 gives access to valuable information as to the speed and efficiency of the internal sensory-
123 motor relationship in the human brain. The present study is specifically devoted to the use
124 of this paradigm to study perceptuo-motor interactions in distinct populations, for some of
125 which perceptuo-motor linkage should be altered due to sensory degradation. The first and
126 principal group of interest of this study is constituted by post-lingually deaf or hearing-
127 impaired adults equipped with a cochlear implant (CI). For post-lingually deaf subjects, the
128 perceptuo-motor linkage, acquired during speech acquisition before deafness, is still
129 efficient during deafness as evidenced by the ability to maintain intelligible speech (Perkell
130 et al. 2000). After CI, the sensory-motor coupling has to deal with a new auditory input

131 provided by the CI. The major question addressed in this study consists of exploring through
132 close-shadowing how sensory-motor relationships are reorganized after CI, comparing prior
133 and present sensorial representation received by the patient.

134 The second group of interest in this study is constituted by normal-hearing elderly adults
135 (NHE). Here, the objective was principally to evaluate the consequences of the potential
136 decline of cognitive and language functions together with a drop in sensory and motor
137 accuracy on the efficiency of perceptuo-motor linkage. An additional objective was to
138 compare the results obtained in this elderly population with those of CI patients. This was
139 because a number of post-lingually deaf CI patients are rather old and there was a need to
140 determine if CI patients' results were influenced by the age of this population.

141 **Perceptuo-motor interactions in Cochlear-Implanted subjects**

142 In speech perception, several factors appear to influence performance in CI patients. In
143 Blamey et al. (2012), 2251 CI patients participated in an auditory test, where they had to
144 recognize phonemes, words and sentences. The experimenters reported that the implant
145 age and duration, the age at the onset of severe to profound hearing loss and the duration
146 of deafness influenced speech perception to a certain extent – although inter-subject
147 variability was quite large in this kind of study.

148 Concerning audiovisual speech perception, CI patients, as well as deaf people, generally
149 show improved lip-reading abilities compared to normal-hearing people, and more
150 importantly, the former seem to present better capacities to integrate visual and auditory
151 speech signals (Goh et al. 2001; Kirk et al. 2002; Kaiser et al. 2003; Rouger et al. 2007).

152 Speech production studies show that auditory feedback provided by the CI results in
153 improved control of phonetic targets a few months after implantation (e.g. Langereis et al.,

154 1998; Lane et al., 2005; Ménard et al., 2007). Furthermore, perturbation studies using either
155 a bite-block task that impedes jaw movement (Lane et al. 2005) or a lip-tube task, that
156 hampers lip rounding (Turgeon et al. 2015), assessed compensation abilities in CI subjects
157 with or without auditory feedback. Results showed that post-lingually CI patients could
158 adapt their articulatory trajectory when it was perturbed, to reach their auditory goals and
159 make their pronunciation intelligible, even when the implant was turned off. Such
160 compensatory strategies suggest that perceptuo-motor abilities acquired during speech
161 acquisition in CI subjects are still at work after deafness. In addition, a PET (Positron Emission
162 Tomography) scanning study on visual speech perception in post-lingually deaf CI patients
163 (Rouger et al. 2012) showed that after a short adaptation period with the implant, there was
164 a decrease of the initially abnormal activity in the superior temporal sulcus, a cross-modal
165 brain area, accompanied by a progressive reactivation of frontal premotor speech areas. This
166 suggests that sensorimotor neuroplasticity after CI provides a progressive reactivation of the
167 audio-visuo-motor linkage in CI subjects.

168 **Perceptuo-motor interactions in normal-hearing elderly subjects**

169 Since post-lingually deaf CI patients are often senior adults, we considered it to be
170 relevant to compare them with an NHE population. The auditory abilities of elderly people
171 decline as aging progresses (Gordon-Salant, 1986; Strouse et al., 1998). Specifically, even
172 with close-to-normal auditory thresholds, normal-hearing elderly people have difficulties in
173 adverse listening conditions such as in a noisy environment (Gelfand et al. 1985; Ohde &
174 Abou-Khalil 2001; Fullgrabe et al. 2013). However, it is still unclear how cognitive decline
175 might influence speech perception. Some studies showed a rather small influence of age-
176 related decline on speech perception in noise (Cienkovski 2002; Cienkovski & Vasil 2010),

177 although a general trend seems to exist of correlation between working memory and speech
178 intelligibility in noise for elderly people (see a review in Füllgrabe & Rosen, 2016)

179 Concerning audiovisual speech perception, elderly people seem to use the benefit of
180 audiovisual presentation during the perception of disturbed signals, like in a noisy
181 environment, at least as well as young adults (Sommers et al. 2005; Sekiyama et al. 2014).

182 Very few studies have investigated speech perceptuo-motor relationships in elderly
183 populations. A behavioral study demonstrated that, similarly to younger adults, elderly
184 people adapt their production in the case of degraded auditory feedback (Liu et al. 2010,
185 2011). Sensorimotor neuroplasticity was also observed in elderly people, linked to age-
186 dependent intelligibility effects mainly found in auditory and motor cortical areas of the
187 brain (Tremblay et al. 2013; Bilodeau-Mercure et al. 2015). Taken together, these behavioral
188 and neuroimaging studies suggest that perceptuo-motor coupling in speech is still active and
189 does not seem severely impaired during ageing.

190 **Objectives of the present study**

191 In the present study, we propose to investigate perceptuo-motor relationships in CI
192 patients and in elderly people, using an auditory and audiovisual close-shadowing paradigm.
193 Based on a previous study (Scarbel et al. 2014), we compared RTs and accuracy to auditory
194 and audiovisual speech stimuli from manual and oral responses.

195 From these results, we hypothesized that CI patients might present overall reduced
196 perceptual performances in the close-shadowing task, particularly in the auditory mode. Our
197 main question was whether a possible replication of faster oral answers compared to
198 manual ones existed in these participants. Such a result would suggest persistence, or
199 restoration, of perceptuo-motor links in these subjects.

200 We analyzed elderly participants with regard to previous studies showing a deficit of
201 speech perception, despite correct auditory abilities. We anticipated obtaining lower
202 performances for our group of elderly participants than for the NHY participants in terms of
203 accuracy or response speed. Here again, as for CI subjects, our principal question concerned
204 the maintenance of faster manual than oral responses.

205

206

207

208

MATERIALS AND METHODS**209 Participants**

210 Four groups of participants performed the experiment. The first group consisted of ten CI
211 participants (3 females and 7 males, mean age: 59 years, range: 27-76). As indicated in Table
212 1, members of the CI group differed in several respects: the age at the onset of deafness
213 varied from 7 to 65 years, the duration of deafness varied from 1 month to 58 years and the
214 duration of CI experience varied from 1 month to 9 years. In addition, seven of the ten
215 participants wore a hearing aid in the non-implanted ear and one participant was bilaterally
216 implanted. The second group consisted of ten NHE participants (4 females and 6 males,
217 mean age: 69 years, range: 63-78). The third group consisted of fifteen NHY participants (10
218 females and 5 males, mean age: 30 years, range: 20-40). In addition, a fourth group of 14
219 NHY participants (11 females and 3 males; mean age: 24 years, range: 19-34) were tested
220 using the same experimental protocol except that auditory stimuli were presented in the
221 existence of background noise (NHY-noise). All normal-hearing participants had normal or
222 corrected-to-normal vision and reported no history of speaking, hearing or motor disorders.
223 The experiment was performed in agreement with the ethical standards laid down in the
224 1964 Declaration of Helsinki and was validated by the CERNI (Local Ethical Comity for Non-
225 Interventional Research).

226

Insert Table 1 about here

227 Results from the two groups of NHY participants were already described in a previous
228 paper (Scarbel et al. 2014). They are presented here because the aim of the present study
229 was to compare results from normal-hearing young adults to post-lingually deaf CI patients
230 and to normal-hearing elderly participants.

231 **Stimuli**

232 Multiple utterances of /apa/, /ata/ and /aka/ VCV syllables were individually produced in
233 a sound-attenuated room by a male French native speaker with no known hearing loss (who
234 did not participate in the perceptive experiment). The three syllables were selected
235 according to the distinct place of articulation of the consonant (stop bilabial /p/, alveolar /t/
236 and velar /k/) and to ensure a gradient of visual recognition between these syllables, notably
237 with the bilabial /p/ consonant known to be more visually salient than alveolar /t/ and velar
238 /k/ consonants). The syllables were audiovisually recorded using an AKG 1000S microphone
239 and a high-quality digital video camera zooming the speaker's face.

240 The corpus was recorded to obtain 4 different occurrences of /apa/, /ata/ and /aka/ with
241 various durations of the initial /a/ vowel (0.5s, 1s, 1.5s and 2s). This was done to present
242 participants with syllables in which the onset of the consonant to be categorized would
243 occur at an unpredictable temporal position. For this, the speaker was asked to maintain the
244 production of the initial vowel while expecting a visual “go” signal. The “go” signal was timed
245 to ensure the four different durations of the initial vowel. The speaker produced 48 stimuli
246 (4 initial durations x 3 types of syllable x 4 repetitions). One distinct utterance was selected
247 for each syllable and for each initial vowel duration so as to obtain 12 distinct stimuli (4
248 initial durations x 3 types of syllables). To remove potential irrelevant acoustic differences
249 between the stimuli, the occurrences of /apa/, /ata/ and /aka/ for an expected given initial
250 duration were then cut at their onset to equalize duration of the first vowel. Similarly,
251 duration of the final vowel was equalized at 240ms for all 12 stimuli.

252 The audio tracks of the stimuli were sampled at 44.1 kHz and presented without noise for
253 CI, NHE and NHY participants. For NHY-noise participants, the 12 stimuli were mixed with

254 white noise, low pass filtered at -6dB /oct, with a signal to noise ratio at -3dB (the signal
255 energy being defined from burst onset to the end of the vowel). In the audiovisual modality
256 of the experiment, the video stream consisted of 572-by-520 pixel/images presented at a 50
257 Hz rate with the speaker's full face presented with blue lips to exaggerate movements of the
258 labial dynamics (Lallouache 1990).

259 **Experimental Procedure**

260 The experiment consisted of two categorization tasks (see Figure 1): a close-shadowing
261 task, where the responses were provided orally, and a manual decision task, where the
262 replies were projected manually. Participants were told that they would be presented with
263 /apa/, /ata/ or /aka/ syllables, displayed either in an auditory or an audiovisual fashion. In
264 the close-shadowing task they were instructed to repeat each syllable as quickly as possible.
265 To do so, they were asked to shadow the initial /a/ vowel and, when the stimulus changed to
266 a consonant, to immediately categorize and repeat the perceived CV syllable (/pa/, /ta/ or
267 /ka/). In the manual decision task, participants were instructed to categorize each syllable by
268 pressing as quickly as possible with their dominant hand one of three keys respectively
269 corresponding to /apa/, /ata/ or /aka/. For each task (oral vs. manual response) and each
270 mode (auditory vs. audiovisual), 16 distinct occurrences of /apa/, /ata/ and /aka/ syllables
271 were presented in a fully randomized sequence of 48 trials. The order of the task, the
272 modality of presentation and the key were fully counterbalanced alongside all the
273 participants.

274 **Please insert Figure 1 here**

275 All groups performed the experiment in a soundproofed room. Participants sat in front of
276 a computer monitor at a distance of approximately 50 cm. The acoustic stimuli were

277 presented at a comfortable sound level, the same one for all participants (around 60 dB
278 Sound Pressure Level). For CI, NHE and NHY participants, the presentation of acoustic stimuli
279 was using a loudspeaker, whereas the presentation of acoustic stimuli was with earphones
280 for NHY-noise participants. This procedure was required because noisy stimuli made acoustic
281 analyses complex and possibly inaccurate if the stimulus and oral feedback were mixed. The
282 presentation software (Neurobehavioral Systems, Albany, CA) was used to control stimulus
283 presentation and to record key answers in the manual task. All participants' productions
284 were recorded using an AKG 1000S microphone for off-line analyses, with a system that
285 ensured synchrony between the stimulus presented to the participant and the participant's
286 response. A brief training session preceded each task. The total duration of the experiment
287 was around 30 min.

288 **Acoustic Analyses**

289 To calculate RTs and the percentage of correct responses in the speech-shadowing task,
290 acoustic analyses of participants' productions were performed using Praat software
291 (Boersma & Weenink 2013). A semi-automatic procedure was first performed to segment
292 these productions. Based on minimal duration and low intensity energy parameters, the
293 procedure involved the automatic segmentation of each utterance depending on the
294 detection of an intensity and duration algorithm. Then, for each presented stimulus,
295 whatever the modality of presentation and response, an experimenter coded the
296 participant's answer and assessed whether it was correct or not.

297 RTs were determined from the burst onset of the stop consonant to categorize the onset
298 of the response (see Figure 1). In the manual decision task, the onset of the response
299 corresponded to the key press. In the close-shadowing task, the burst onset of the stop

300 consonant uttered by the participant in response to the stimulus was determined by
301 inspecting the acoustic waveform and spectrogram information. RTs (reaction times) were
302 only computed for correct answers: omissions or any type of errors (replacing one
303 consonant by another or producing two consonants or two syllables in the close-shadowing
304 task) were removed from the analyses.

305 **Further response analyses**

306 For each group, the percentage of correct responses and median RT were individually
307 determined for each participant, each task, each modality and each syllable. Further
308 statistical analyses were carried out independently on a percentage of correct answers and
309 on median RTs. Firstly, repeated-measure ANOVAs were performed separately for the NHE
310 and CI groups, with the task (close-shadowing against manual decision), the modality
311 (auditory vs. audiovisual) and the syllable (/apa/, /ata/ or /aka/) as within-subjects variables.
312 Secondly, comparisons between the four groups were obtained by performing two other
313 repeated-measure ANOVAs on RTs and the percentages of correct responses with the CI,
314 NHE, NHY, NHY-noise groups as between-subject variables and the tasks (close-shadowing or
315 manual decision), modality (auditory or audiovisual) and the syllable (/apa/, /ata/, /aka/) as
316 within-subject variables. For all the following analyses, the significance level was set at $p =$
317 0.05 and Greenhouse–Geisser corrected (in the case of violation of the sphericity
318 assumption) when appropriate. All comparisons reported refer to posthoc analyses
319 conducted with Bonferroni tests.

320 Then, to assess the relationship between accuracy and reaction time, an analysis of
321 correlation between reaction time and percentage of correct answers was carried out for

322 each group and each condition separately (oral and manual responses for audio and
323 audiovisual stimuli).

324 Finally, since the CI group appears to be highly variable in terms of subject age, age of
325 deafness onset, duration of deafness and duration of CI experience (see Table 1), individual
326 ANOVAS on RT values were performed for each CI participant, with the task (oral vs. manual
327 response) and the modality (auditory vs. audiovisual presentation) as independent variables.
328 This was done to assess the effects of these variables on RTs for each CI subject.

329

330

RESULTS

331 The results of the four groups are shown in Figure 2 (for reaction times) and Figure 3 (for
332 percentages of correct responses).

333

Insert Figures 2 & 3 here

334

Normal-hearing elderly participants

336 Concerning reaction times, the main effect of task was found to be significant
337 ($F(1,9)=255.31$; $p<.001$), with faster RTs for oral responses than for manual ones (on
338 average, oral responses: 238 ms; manual responses: 568 ms). In addition, the main effect of
339 syllable was found to be significant ($F(1,9)=12.67$; $p<.001$) with faster responses for /apa/
340 (374 ms) than for /aka/ (429 ms) and faster responses for /aka/ than for /ata/ (451 ms).
341 There was no effect of modality, either alone or in interaction.

342 Concerning the percentages of correct responses, the main effect of task was significant
343 ($F(1,9)=25.5$; $p<.001$) with a lower value for oral than for manual responses (on average,
344 oral responses: 81%; manual responses 98%). In addition, the main effect of syllable was
345 found to be significant ($F(2,18)=7.9$; $p<.005$), with a lower percentage of correct answers for
346 /apa/ (98%) than for /ata/ (81%). Modality did not provide any significant effect, alone or in
347 interaction. Finally, the interaction between task and syllable was found to be significant
348 ($F(2,18)=10.08$; $p<.005$). However, significant differences between syllables were only found
349 for the oral task (higher percentage for /apa/ than for /aka/ or /ata/, and higher for /aka/
350 than for /ata/).

Cochlear implanted participants

352 Concerning reaction times, the main effect of task was significant ($F(1,9)=36.27$; $p<.001$)
353 with much faster RTs for oral responses than for manual ones (on average, oral responses:

354 469 ms; manual responses: 826 ms). The main effect of modality was also significant
355 ($F(1,9)=7.10$; $p=.02$) with faster responses for audiovisual stimuli than for audio stimuli (on
356 average, audiovisual stimuli: 626 ms; audio stimuli: 669ms). Additionally, the main effect of
357 syllable was found to be significant ($F(2,18)=17.39$; $p<.001$), with faster responses for /apa/
358 and /aka/ than for /ata/ (/apa/: 586 ms; /aka/: 621 ms; /ata/: 736 ms). Finally, a significant
359 interaction between modality and syllable was found ($F(2,18)=6.05$; $p<.01$) showing that the
360 advantage of audiovisual presentation was present solely for /ata/ (audio: 782 ms;
361 audiovisual: 690 ms).

362 Concerning the percentages of correct responses, no significant difference was found
363 either between audio (69%) and audiovisual (79%) stimuli, or between oral (67%) and
364 manual (78%) answers. However, the main effect of syllable was significant ($F(2,18)= 14.45$;
365 $p<.001$), with a higher percentage of correct responses for /apa/ and /aka/ than with /ata/ (
366 /apa/: 86%; /aka/: 88%; /ata/: 44%). No interactions were significant, neither between
367 modality and task, nor between syllable and modality or task.

368 **Comparison between the four groups**

369 **Reaction times (see Figure 2)**

370 The main effect of group was significant ($F(3,45)=24.43$; $p<.001$), with slower reaction
371 times for CI participants than for the three normal-hearing groups. No significant difference
372 was observed between elderly participants and the two groups of young participants.
373 However, NHY participants were faster than NHY-noise participants (overall, on average, CI:
374 648 ms; NHY-noise: 484 ms; NHE: 418 ms; NHY: 351 ms). The main effect of task
375 ($F(1,45)=263.95$; $p<.001$) and of modality ($F(1,45)=21.09$; $p<.001$) were also found to be
376 significant. For the task, oral responses were faster than manual ones (on average, 313 ms

377 vs. 600 ms). For the modality, RTs were faster in the audiovisual modality than in the
378 auditory modality (on average, 446 ms vs. 467 ms).

379 A significant interaction between group and modality ($F(3,45)=6.93$; $p<.001$) further
380 indicated that faster RTs were observed in the audiovisual modality than in the auditory
381 modality, only for CI participants (on average, 626 ms vs. 669 ms) and for NHY-noise
382 participants (on average, 461 ms vs. 507 ms). No difference was found for NHE participants
383 (on average, 414 ms vs. 422 ms), or for NHY participants (on average, 354 ms vs. 349 ms).

384 The interaction between modality and syllable was also significant ($F(2,90)=8.94$; $p<.001$),
385 showing that irrespective of the group, faster RTs in the audiovisual modality compared to
386 the auditory modality only occurred for the /apa/ syllable.

387 In addition, there was a significant interaction between group and syllable ($F(6,90)=3.77$;
388 $p<.005$). For NHY-noise participants, RTs for /apa/ were faster than for /ata/ and /aka/, for
389 NHE participants, RTs for /apa/ were faster than for /ata/, whereas for CI patients, RTs for
390 /apa/ and /aka/ were faster than for /ata/.

391 Finally, a significant interaction between group, modality and syllable ($F(6,90)=4.04$;
392 $p<.005$) was found, showing that the modality effect was only present for the /apa/ syllable
393 for the NHY-noise group.

394 **Percentage of correct responses (see Figure 3)**

395 The main effect of group was significant ($F(3,45)=57.66$; $p<.001$), with a higher percentage
396 of correct responses for NHY participants (95%) and NHE participants (89%) than for CI
397 participants (72%) and a higher percentage of correct answers for CI participants than for
398 NHY-noise participants (60%). There was no significant difference between NHY and NHE
399 participants.

400 The main effect of task was also significant ($F(1,45)=71.36$; $p<.001$), with a lower
401 percentage of correct responses for oral responses than for manual responses (73% vs.
402 86%). The main effect of modality was also significant ($F(1,45)=28.76$; $p<.001$), with more
403 accurate answers for audiovisual than for auditory stimuli (82% vs. 77%).

404 A significant interaction between group and task was observed ($F(3,45)=6.26$; $p<.005$),
405 indicating that the difference between oral and manual responses was only present for NHE
406 participants (81% vs. 98%) and NHY-noise participants (50% vs. 71%). Moreover, a significant
407 interaction between group and modality ($F(3,45)=12.97$; $p<.001$) showed that the modality
408 benefit was only present for NHY-noise participants (audiovisual: 68%; audio: 53%).

409 As for RTs, the interaction between modality and syllable was significant ($F(2,90)=11.05$;
410 $p<.001$), showing that the effect of the audiovisual modality was present only for the /apa/
411 syllable.

412 The interaction between group and syllable was also significant ($F(6,90)=8.41$; $p<.001$),
413 with no difference observed between the three syllables for NHE and NHY participants
414 whereas for NHY-noise participants, /apa/ was more clearly recognized than /ata/ and, for CI
415 participants, /apa/ and /aka/ were recognized better than /ata/.

416 In addition, an interaction between task and syllable was observed ($F(2,90)=19.96$;
417 $p<.001$), showing that for oral responses, /apa/ was recognized better than /aka/, which was
418 of improved recognition compared to /ata/. In contrast, for manual responses, no difference
419 was observed between /apa/ and /aka/ syllables but these were better recognized than
420 /ata/.

421 The interaction between group, modality and syllable was found to be significant
422 ($F(6,90)=4.72$; $p<.001$), showing that for NHY-noise participants, the effect of modality was
423 present only for /apa/ (on average, 92%; audio: 61%).

424 Finally, the interaction between group, task and syllable was significant ($F(6,90)=5.07$;
425 $p<.001$). Focusing on significant effects, for NHE participants, /apa/ was recognized better
426 than /ata/ and /aka/ with oral responses; for CI participants, /apa/ and /aka/ were
427 recognized better than /ata/ for both oral and manual responses; and for NHY-noise, /apa/
428 was recognized better than /ata/ and /aka/ for oral responses, whereas /apa/ and /aka/
429 were recognized better than /ata/ for manual responses.

430 **Correlation between reaction times and percentages of correct responses (see Figure 4)**

431 For the NHE and CI groups, the correlation between RTs and the percentage of correct
432 answers was analyzed for each group and each condition separately (oral and manual
433 responses for audio and audiovisual stimuli). For NHE participants (Figure 4a, b), all
434 correlations show the same tendency to be positive, although significance is only observed
435 for oral responses to audiovisual stimuli ($r^2=0.6$, $p<.01$). This shows that faster participants
436 tend to obtain less accurate responses, in agreement with classical observations on the
437 speed-accuracy tradeoff. Importantly, Figure 4 shows clearly that the difference between
438 oral and manual responses appears to be independent of this tradeoff. Indeed, even in
439 regions where accuracy is similar, oral responses are much faster than manual ones.

440 In the same figure, we superimposed data for the NHY group without noise onto that of
441 the NHE group. It seems likely that the two groups behave differently, younger people being
442 globally more accurate and faster than their elders (mean values for NHE: 89% and 418 ms;
443 mean values for NHY: 95% and 351 ms). However, this was not displayed in the ANOVA,

444 where no significant differences appeared between the two groups for either measurement,
445 probably because of too small a number of subjects.

446 For CI patients, the portrait is more complex. Indeed, the only significant correlation is
447 negative, for manual responses to audio stimuli ($r^2=.45$, $p<.05$). Such a negative correlation is
448 at odds with the speed-accuracy tradeoff, which is quite a general process (Fitts, 1966;
449 Wickelgren, 1977), well described in computational decision models (Ratcliff & McKoon,
450 2008). This confirms the high degree of inter-individual differences between CI subjects.
451 Indeed, the negative correlation shows that some subjects are both more accurate and rapid
452 than others, hence that they have recovered a higher level of speech decoding.

453 **Insert Figure 4 here**

454 **Individual analysis in the CI group (see Table 2)**

455 Because of this high degree of inter-individual variability in the CI group, the last analysis we
456 did aimed to assess whether the oral-manual difference in response times could be found in
457 each individual subject. To this end, for each CI participant, a two-factors ANOVA was
458 performed on reaction time values with the tasks (oral, manual) and modality (audio,
459 audiovisual) as independent variables – with 48 estimations per condition, grouping the
460 three syllables.

461 Individual mean values and significance levels for task and modality are shown in Table 2.
462 Importantly, for all participants the main effect of task was highly significant (with p values
463 inferior to .001), with faster oral than manual responses. Additionally, four participants
464 obtained a significant modality effect with faster responses to audiovisual than to audio
465 stimuli. Interaction between task and modality was significant for two participants (with p

466 values inferior to .001), with significantly larger reaction times for audio than for audiovisual
467 stimuli in oral responses.

468

Insert Table 2 here

469

470

DISCUSSION

471 The goal of this study was to assess the perceptuo-motor link in post-lingually deaf CI
472 patients and in NHE participants, through an auditory and audiovisual close-shadowing
473 experiment. Results of these two groups of patients were compared with results from a
474 previous study (Scarbel et al., 2014) of two groups of normal-hearing young participants,
475 with or without the addition of acoustic noise. In the previous study, we obtained faster RTs
476 in the oral than in the manual task; and more surprisingly, less accurate oral than manual
477 responses in noise. Additionally, concerning the modality of presentation, responses to
478 audiovisual stimuli were found to be faster and more accurate than audio stimuli in noise,
479 though not in silence.

480 Basically, we repeated this global pattern for the two groups under study here, although
481 with some differences that will be discussed in detail below. These differences include a
482 global trend for a slight degradation in performance for older NH subjects, with a tendency
483 to display slower RTs and a lower percentage of correct answers than younger participants
484 without noise. More importantly, there is a significant degradation of performances in CI
485 patients who presented slower replies and a lower percentage of correct answers than those
486 of the three normal-hearing groups. That is, young participants with or without noise, and
487 normal-hearing elderly participants. However, of these three the young participants in noise
488 produced the most errors.

489 We will now discuss in more detail the pattern of results related to the role of task, and
490 modality of presentation.

491 **Effect of task: manual vs. oral responses**

492 **Normal-hearing elderly participants**

493 Elderly participants present response patterns that are quite similar to those of young
494 participants with noise, concerning the effect of the task. Indeed, we obtain for elderly
495 participants both faster reaction times and a lower percentage of correct answers for oral
496 than for manual responses. The difference in speed between oral and manual responses was
497 already seen in young adults with clear acoustic stimuli. The fact that there is also a
498 significant difference in accuracy in more elderly subjects, even with clear acoustic stimuli, is
499 likely due to the slight degradation in accuracy – as in reaction times – with age. These data
500 allow us to confirm the results obtained in our previous study (Scarbel et al., 2014).

501 **Cochlear-implanted participants**

502 Crucially, as for normal-hearing groups, oral responses in CI participants were found to be
503 faster than manual responses – actually much faster, more than 350 ms in average.
504 Importantly, the difference between reaction times in the two tasks appears to be significant
505 for all CI patients, even subjects recorded just after implantation (1 to 3 months).

506 However, in contrast to normal-hearing young participants with noise and elderly
507 participants, no significant difference between oral and manual answers in terms of
508 response accuracy was observed for CI patients. This lack of significant difference (in spite of
509 a 10% trend) is probably due to the variability of performances for the CI group, together
510 with the small size of the group.

511 **Effect of the modality**

512 **Normal-hearing elderly participants**

513 Regarding modality, it is important to notice that elderly participants presented response
514 patterns that were quite similar to those of young participants without noise. Elderly
515 participants did not present a benefit of the visual modality, since no difference in reaction

516 times and in the percentage of correct answers were observed between responses for
517 auditory and audiovisual stimuli. This lack of visual modality benefit might appear surprising,
518 considering the relatively low accuracy scores in the auditory condition (around 80% globally
519 in the oral response mode) since most studies on audiovisual perception in elderly people
520 conclude that audiovisual integration abilities are quite similar in elderly participants as in
521 younger ones (Cienkowski et al. 2002; Sommers et al. 2005, Stevenson et al. 2015), if not
522 better (Sekiyama et al. 2014). This might suggest that audiovisual integration is less efficient
523 in elder subjects under time pressure.

524 **Cochlear-implanted participants**

525 Concerning stimuli modality presentation, CI patients, like normal-hearing young
526 participants with noise, obtained faster responses for audiovisual stimuli than for audio
527 stimuli. However, whereas normal-hearing young participants with noise obtained higher
528 percentages of correct answers for audiovisual stimuli, no difference between the two
529 modalities was observed for CI patients at the group level – although an effect was displayed
530 individually for 4 subjects. These results appear only partly in agreement with those of
531 previous studies exploring lip-reading abilities in cochlear-implanted patients (Goh et al.
532 2001; Kirk et al. 2002; Bergeson et al. 2003; Kaiser et al. 2003; Rouger et al. 2012), which
533 showed a major role of vision and suggested a good, if not better than normal, capacity to
534 integrate visual and auditory abilities for these participants. Indeed, it is somewhat puzzling
535 that CI subjects did not benefit much from the visual modality in terms of accuracy,
536 considering the relatively low accuracy scores in the auditory modality. This is perhaps due
537 to the specificity of the task with respect to time pressure, which could have perturbed the
538 audiovisual fusion process.

561 Normal-hearing elderly participants display similar trends, with a difference both in speed
562 and accuracy between oral and manual responses, which is in agreement with the analysis-
563 by-synthesis model shown in Figure 5. The similarity between younger and elder participants
564 and the significantly quicker oral responses would indicate in this framework that the
565 perceptuo-motor loop is still effective in the speech perception process in elderly
566 participants, as suggested by Liu et al. (2010, 2011).

567 Concerning CI patients, the existence of significant differences in response times between
568 oral and manual responses, together with the global increase in response time in both
569 response modalities, can once again be interpreted in the framework of the model proposed
570 by Skipper et al. (2007) (Figure 5). In this case, we might assume that multisensory
571 processing is degraded and delayed because of auditory difficulties (Stage 1'). Then the
572 whole process would basically be preserved – though probably degraded – as in normal-
573 hearing subjects. Therefore, globally, all responses would be slower and less accurate, but
574 the difference between oral and manual responses would remain the same as in normal-
575 hearing participants.

576 **Close-shadowing, an interesting tool for assessing the recovery of the perceptuo-motor**
577 **connection after cochlear implantation**

578 These results allow us to shed light on the perceptuo-motor linkage restoration in
579 cochlear-implanted participants with a close-shadowing paradigm, resulting in faster but
580 possibly less accurate responses in the speech task than in the manual task. Indeed,
581 whatever the interpretation for the speed advantage in the oral task, it shows that cochlear-
582 implanted subjects have recovered an efficient auditory-to-articulatory or motor connection.
583 This connection enables them to efficiently convert the auditory input provided by the

584 implant into the appropriate motor command. Strikingly, this ability is displayed soon after
585 implantation, even after less than 3 months. Crucially, this restoration seems to be efficient
586 in spite of the incomplete recovery of auditory perception abilities, since CI patients
587 presented lower scores than normal-hearing participants.

588 Apart from this result about the (at least partial) recovery of the perceptuo-motor link in
589 CI subjects, the present study provides a major methodological output. Indeed, it appears
590 that close-shadowing might constitute an interesting and useful paradigm to evaluate
591 sensory-motor interaction in CI patients after implantation. We propose that this paradigm
592 could constitute a new audiological test for the rehabilitation toolkit for CI patients or
593 hearing-impaired patients, and eventually for patients with other types of speech
594 impediment.

595 **Conclusion**

596 The results of the present study suggest that oral responses to auditory or audiovisual
597 syllables are faster than manual responses in normal-hearing adults, both young and elderly,
598 and in CI patients. Results for elder normal-hearing participants show that the perceptuo-
599 motor coupling is still efficient and rather stable in these subjects. Importantly, the gain in
600 speed of oral responses, maintained for CI patients in spite of a global degradation in
601 performance, suggests that the perceptuo-motor coupling has been recovered, at least in
602 part, even following short periods of implantation. We suggest an interpretation of these
603 results as a new illustration of the perceptuo-motor coupling in the framework of speech
604 perception theories, suggesting that oral responses would be generated at an early stage of
605 a sensorimotor loop, whereas manual responses would be formulated at the end of this
606 loop, leading to slower but more accurate responses.

ACKNOWLEDGEMENTS

This work was supported by the French National Research Agency (ANR) through funding from the PlasMody project (Plasticity and Multimodality in Oral Communication for the Deaf). It also received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 339152 - "Speech Unit(e)s").

REFERENCES

- Bergeson, T. R., Pisoni, D. B., Davis, R. A. O. (2003). A Longitudinal Study of Audiovisual Speech Perception by Children with Hearing Loss Who have Cochlear Implants. *The Volta review*, 103(4), 347-370.
- Bilodeau-Mercure, M., Ouellet, C., Tremblay, P. (2015). Vieillissement oro-facial et production de la parole. Poster presented at the *Congrès québécois de recherche en adaptation-réadaptation*. Boucherville, Qc, Canada.
- Blamey, P., Artieres, F., Baskent, D., et al. (2012). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implant: an update with 2251 patients. *Audiology & Neurology* 18, 36-47.
- Blumstein, S. E. & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66, 1001-1017
- Boersma, P. (2001). Praat, a system for doing phonetics by computer, *Glott International*, 5(9/10), 341–345.
- Cienkowski, K. M. & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing*, 23(5), 439-449.
- Cienkowski, K. M., & Vasil-Dilaj, K. (2010). The Role of Central Auditory and Cognitive Processing Ability in Speech Perception for Elderly Listeners. *Perspectives on Hearing and Hearing Disorders Research and Diagnostics*, 14(2), 30.

- Dias, J.W., Rosenblum, L.D. (2016). Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception & Psychophysics*, 78(1), 317-33.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.
- Fitts, P. M. (1966). Cognitive aspects of information processing. 3. Set for speed versus accuracy. *Journal of Experimental Psychology*, 71(6), 849-57.
- Fowler, C. A. (1986). An event approach to the study of speech perception for a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. A., Brown, J. M., Sabadini, L., et al. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49, 296–314.
- Füllgrabe, C. (2013). Age-Dependent Changes in Temporal-Fine-Structure Processing in the Absence of Peripheral Hearing Loss. *American Journal of Audiology*, 22(2), 313.
- Füllgrabe, C. & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in psychology*, 7, 1268
- Galantucci, B., Fowler, C. A., Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13, 361-377.
- Gelfand, S. A., Piper, N., Silman, S. (1985). Consonant recognition in quiet as a function of aging among normal hearing subjects. *The Journal of the Acoustical Society of America*, 78(4), 1198-1206.

- Goh, W. D., Pisoni, D. B., Kirk, K. I., et al. (2001). Audio-Visual Perception of Sinewave Speech in an Adult Cochlear Implant User: A Case Study. *Ear and Hear*, 22(5), 412-419.
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-79.
- Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *The Journal of the Acoustical Society of America*, 80(6), 1599-607.
- Guenther, F. H. (2006). Cortical organisations underlying the production of speech sounds, *Journal of Communication Disorders*, 39, 350–365.
- Hickock, G. & Poeppel, D. (2007). The cortical organization of speech processing, *Nature Reviews Neuroscience*, 8, 393-402.
- Kaiser, A. R., Kirk, K. I., Lachs, L., et al. (2003). Talker and Lexical Effects on Audiovisual Word Recognition by Adults With Cochlear Implants. *Journal of speech, language, and hearing research : Journal of Speech, Language, and Hearing Research*, 46(2), 390-404.
- Kirk, K. I., Pisoni, D. B., Lachs, L. (2002). Audiovisual integration of speech by children and adults with cochlear implants. *Proceedings : ICSLP. International Conference on Spoken Language Processing, 2002*, 1689-1692.
- Lallouache, M.T. (1990) Un poste « visage-parole ». Acquisition et traitement des contours labiaux. (A 'face-speech' workstation. Acquisition and processing of labial contours). *Proceedings of the eighteen Journées d'Etudes de la Parole, Montreal, QC.*

- Lane, H., Denny, M., Guenther, F. H., et al. (2005). Effects of bite blocks and hearing status on vowel production. *The Journal of the Acoustical Society of America*, *118*(3;1), 1636-1646.
- Langereis, M.C., Bosman, A.J., Van Olphen, A.F., & Smoorenburd, G.F. (1998). Effect of cochlear implantation on voice fundamental frequency in post-lingually deafened adults. *Audiology*, *37*, 219-230.
- Lieberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception Revised. *Cognition*, *21*(1), 1–36.
- Lindblom, B., & Maddieson, I. (1988). Phonetic Universals in Consonant Systems. *Language, Speech and Mind*, 62-78.
- Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (éd.), *Speech Production and Speech Modelling* (p. 403-439). Springer Netherlands.
- Liu, H., Wang, E. Q., Chen, Z., et al. (2010). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. *The Journal of the Acoustical Society of America*, *128*(6), 3739-3746.
- Liu, H., Meshman, M., Behroozmand, R., et al. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *122*(5), 951-957.
- Liu, P., Chen, Z., Jones, J. A., et al. (2011). Auditory Feedback Control of Vocal Pitch during Sustained Vocalization: A Cross-Sectional Study of Adult Aging. *PLoS ONE*, *6*(7), e22791.

- Marslen-Wilson WD (1985). Speech shadowing and speech comprehension. *Speech Communication*, 4: 55-73.
- Ménard, L., Polak, M., Denny, M., et al. (2007). Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *The Journal of the Acoustical Society of America*, 121(6), 3790-3801.
- Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception, & Psychophysics*, 75, 557-575.
- Nye, P. & Fowler, C. A. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31: 63–79.
- Ohde, R. N., & Abou-Khalil, R. (2001). Age differences for stop-consonant and vowel perception in adults. *The Journal of the Acoustical Society of America*, 110(4), 2156-2166.
- Perkell, J.S., Lane, H., Svirsky, M., et al. (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *The Journal of the Acoustical Society of America*, 91, 2961-2979.
- Perkell, J.S., Guenther, F.H., Lane, H., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of phonetics*, 28, 233-272.
- Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. In J. McQueen and A. Cutler (eds) SWAP special issue, *Language and Cognitive Processes* 16(5/6), 691-698.

- Porter, R. & Castellanos, F. (1980). Speech production measures of speech perception: Rapid shadowing of VCV syllables. *The Journal of the Acoustical Society of America* 67, 1349–1356.
- Radcliff, R. & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20:4, 873-922
- Rouger, J., Lagleyre, S., Fraysse, B., et al. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7295-7300.
- Rouger, J., Lagleyre, S., Démonet, J.-F., et al. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Human Brain Mapping*, 33(8), 1929-1940.
- Scarbel, L., Beautemps, D., Schwartz, J.-L., et al. (2014). The shadow of a doubt? Evidence for perceptuo-motor linkage during auditory and audiovisual close-shadowing. *Frontiers in Psychology*, 5(568).
- Schwartz, J.L., Basirat, A., Ménard, L., et al. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25, 336-354.
- Sekiyama, K., Soshi, T., Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology*, 5.
- Skipper, J. I., vanWassenhove, V., Nusbaum, H. C., et al. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17, 2387-2399.

- Skipper, J.I., Devlin, J.T., Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain & Language*, 164, 77-105.
- Sommers, M. S., Tye-Murray, N., Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hear*, 26(3), 263-275.
- Stevens, K.N., and Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.) *Models for the Perception of Speech and Visual Form*. (pp. 88-102). Cambridge, MA: MIT Press.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358-1368.
- Strouse, A., Ashmead, D.H., Ohde, R.N., Grantham, D.W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America*, 104(4), 2385-99.
- Sumbly, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2), 212-215.
- Tremblay, P., Dick, A. S., Small, S. L. (2013). Functional and structural aging of the speech sensorimotor neural system: functional magnetic resonance imaging evidence. *Neurobiology of Aging*, 34(8), 1935-1951.
- Turgeon, C., Prémont, A., Trudeau-Fisette, P., et al. (2015). Exploring production-perception relationships in normal hearing and cochlear implant adults: a lip-tube perturbation study. *Clinical Linguistics and Phonetics*, 29(5), 378-400.

Wickelgren, WA. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67–85.

FIGURE LEGENDS

Figure 1: Experimental design. Reaction times were measured between stimulus and response bursts (plosion release) for oral responses and between stimulus burst and key press for manual responses.

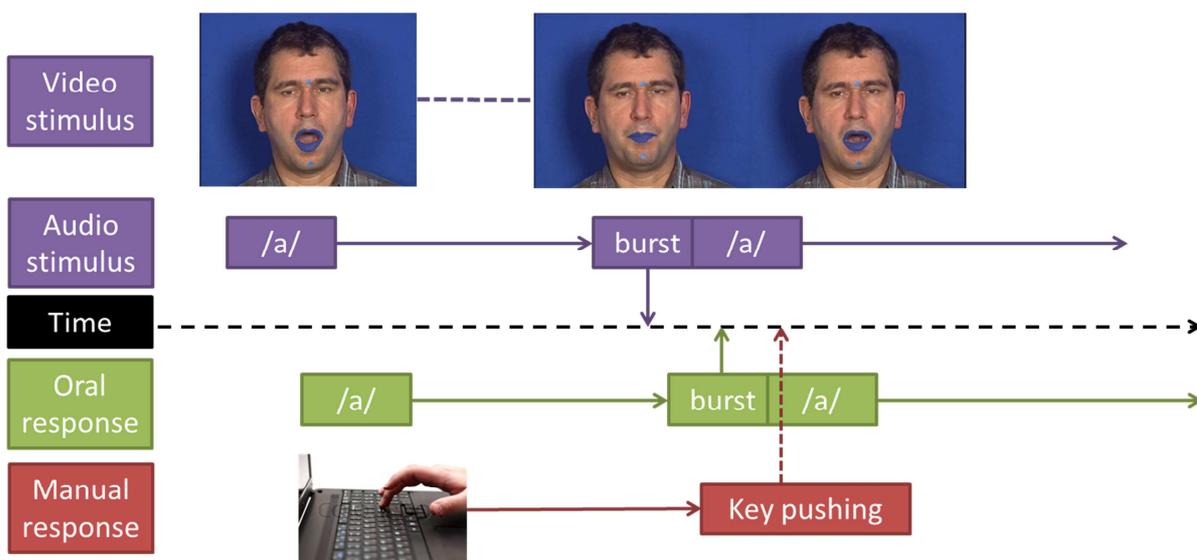


Figure 2: Mean RTs (in ms) in each group (normal-hearing young participants with noise (NHY-noise), normal-hearing young participants without noise (NHY), normal-hearing elderly participants (NHE), cochlear-implanted-participants (CI)), task (oral, manual response) and modality of presentation (auditory, audiovisual). Error bars represent standard errors of the mean.

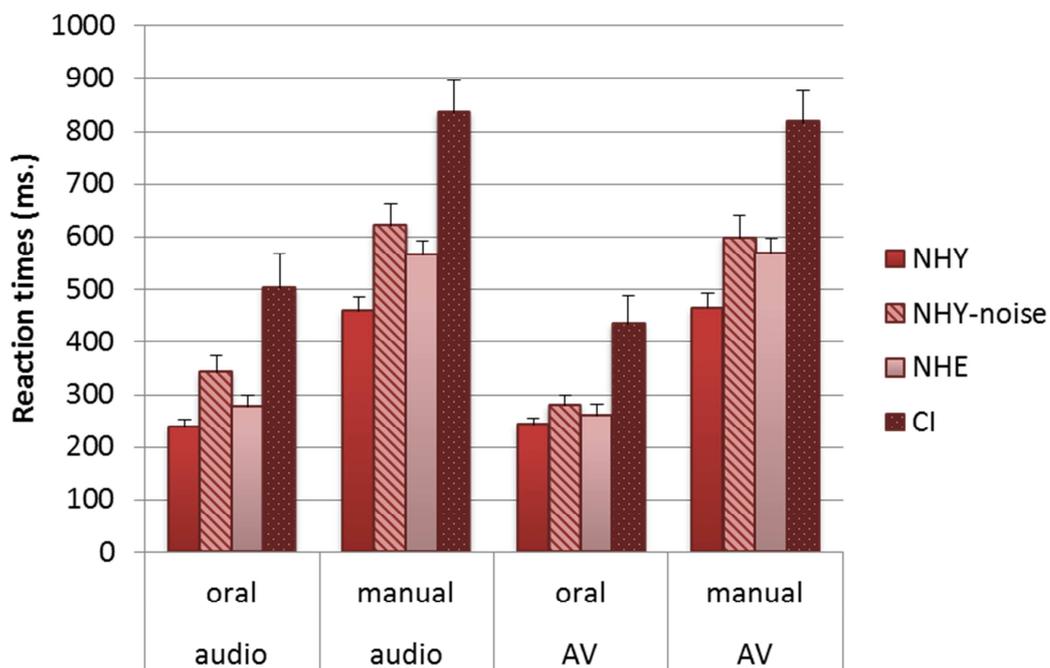


Figure 3: Mean percentage of correct identification in each group (normal-hearing young participants with noise (NHY-noise), normal-hearing young participants without noise (NHY), normal-hearing elderly participants (NHE), cochlear-implanted participants (CI)), task (oral, manual response) and modality of presentation (auditory, audiovisual). Error bars represent standard errors of the mean.

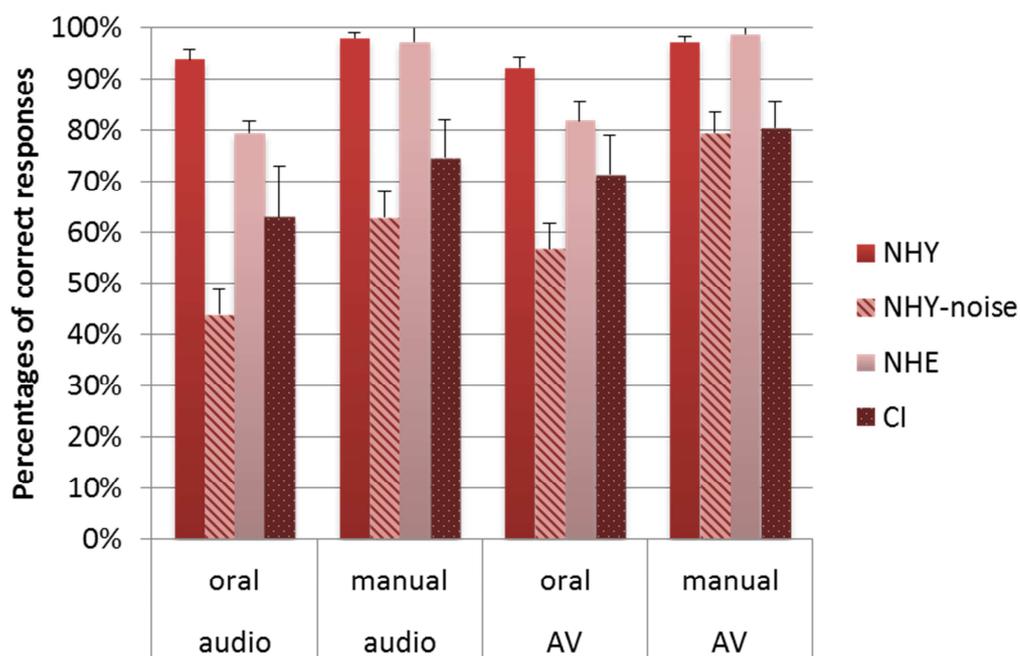


Figure 4: Correlation between RT (in ms) and percentage of correct answers (in %) in the NHE and NHY (A & B) and CI (C & D) groups, in response to audio (A & C) and audiovisual (B & D) stimuli. Correlation was significant for NHE subjects in the oral AV condition, with a positive slope. It was significant for CI subjects in the manual A condition, with a negative slope. To attempt to avoid ceiling effects with response probabilities close to 100%, another analysis was realized using an arcsine transformation, but it provided exactly the same pattern of significant correlations.

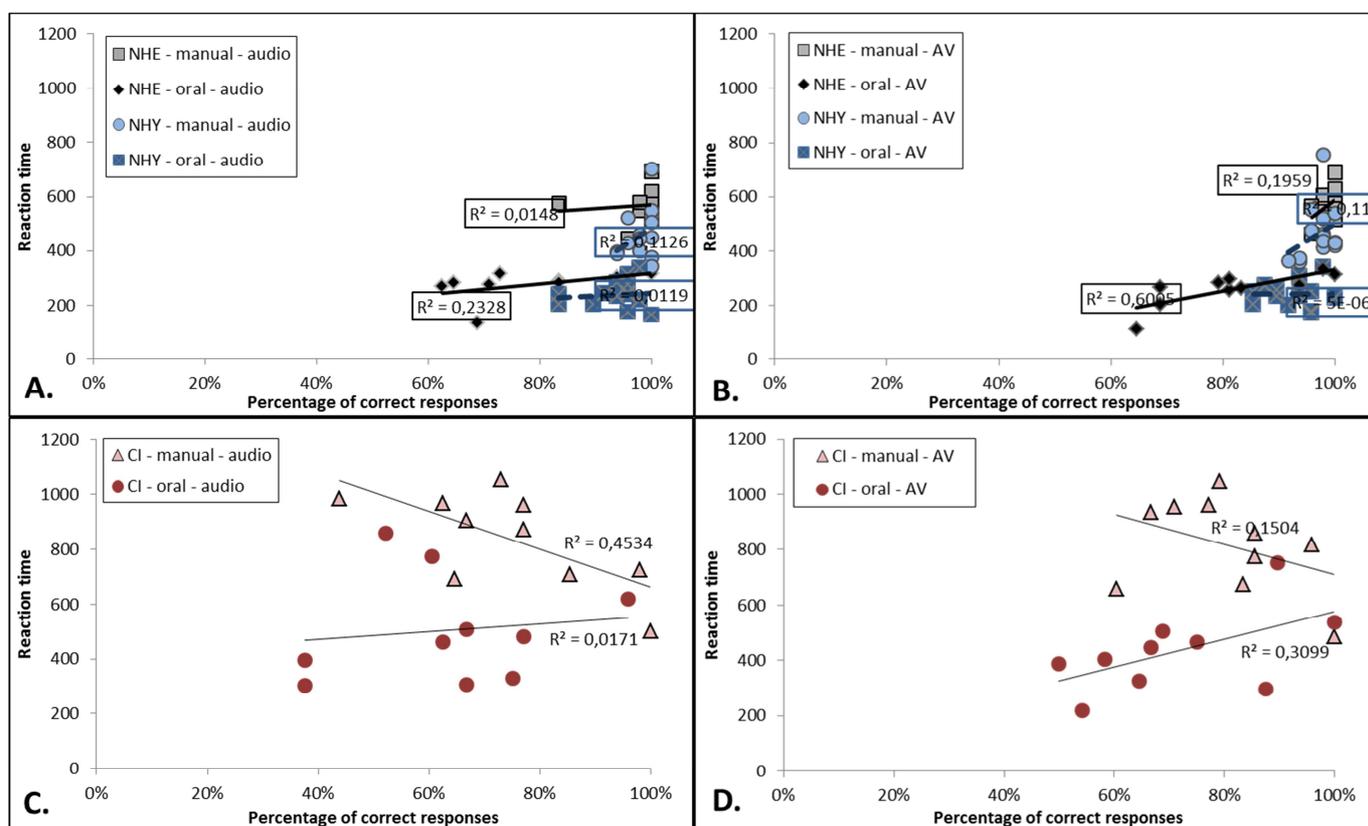
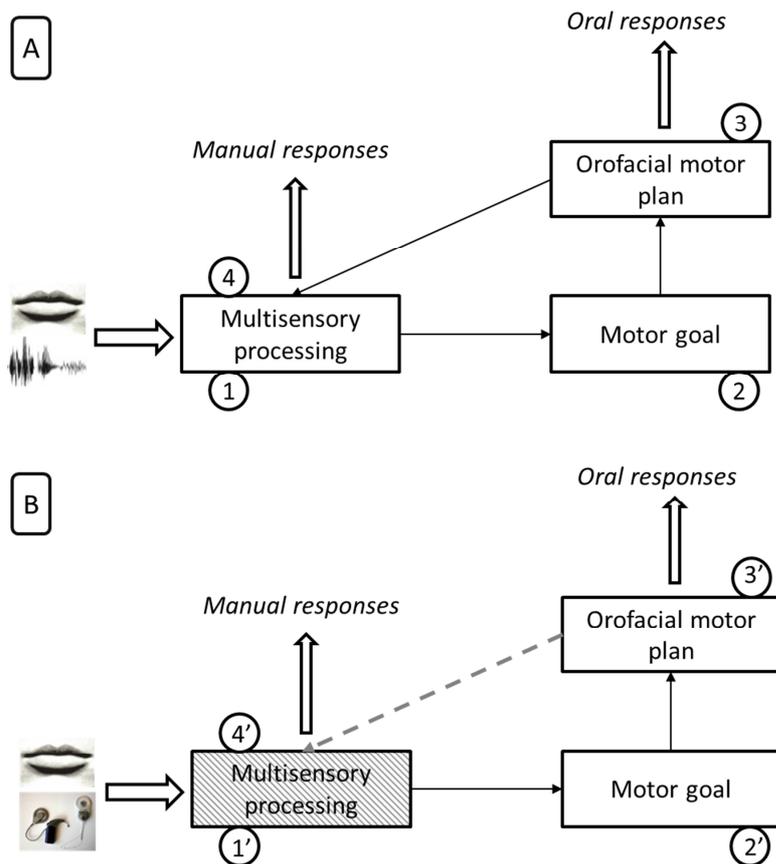


Figure 5: Possible auditory-motor relationships according to manual and oral response modes for normal-hearing participants (A) and cochlear-implanted participants (B) (adapted from Skipper et al.'s model, 2007). Multisensory processing would be degraded in CI patients (hatched on the figure).



TABLES

Table 1: Characteristics of the ten participants with cochlear implants

Sex	Age (in years)	Age at the onset of deafness (in years)	Hearing aid	Duration of deafness	Duration of CI experience
M	65	7	No	58 years	1 month
M	56	20	Yes	35 years	3 months
F	66	32	Yes	25 years	9 years
M	60	59	Yes	1 month	1 years, 4 months
F	43	20	Yes	13 years	2 months
M	27	25	Bilateral CI	2 months	2 years, 6 months
F	67	65	Yes	2 years	7 months
M	72	40	Yes	30 years	5 months
M	76	48	No	27 years	3 years, 4 months
M	57	48	Yes	8 years.	10 months

Table 2: Individual mean RT values and significance levels for modality and task in CI patients

	Task			Modality			Task x Modality
	Oral (ms)	Manual (ms)	Significance	Audio (ms)	Audiovisual (ms)	Significance	Significance
CI1	193	953	p<.001	797	664	p<.001	n.s.
CI2	306	480	p<.001	440	384	p<.001	p<.001
CI3	415	674	p<.001	541	523	n.s.	n.s.
CI4	435	996	p<.001	742	717	n.s.	n.s.
CI5	308	692	p<.001	522	486	p<.005	p<.001
CI6	653	779	p<.001	705	726	n.s.	n.s.
CI7	794	939	p<.001	897	844	n.s.	n.s.
CI8	330	1052	p<.001	830	726	n.s.	n.s.
CI9	653	841	p<.001	840	684	p<.001	n.s.
CI10	420	858	p<.001	680	609	n.s.	n.s.