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

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

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

Lucie Scarbel[1], Denis Beautemps[1], Jean-Luc Schwartz[1] & Marc Sato[2]

[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.

ABSTRACT

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

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

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

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

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

KEY-WORDS

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

INTRODUCTION

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

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

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

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

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

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

Close-shadowing therefore appears to provide “a continuous reflection of the outcome of the process of language comprehension, (and) provides us with uniquely privileged access to the properties of the system” (Marslen-Wilson, 1985, p. 55). However, the interpretation relating the high speed of close-shadowing responses to the articulatory nature of speech percepts e.g. Galantucci et al. (2006) remains controversial. Firstly, it can be suggested that the rapidity of oral responses only reflects a sensory-motor coupling process in which auditory speech decoding would automatically evoke preparatory motor responses. Oral responses would be particularly fast if the task engaged the subjects to answer rapidly, but auditory processing would not necessarily use motor representation in its functioning. This is compatible with the way exemplary models are applied to speech (e.g. Pierrehumbert, 2001) and to how "echos" can be introduced as close-shadowing models in such frameworks (Goldinger, 1998). A second line of controversy is supplied by the observation that close-shadowing involves a certain amount of linguistic processing [related to phonology, syntax and semantics (Marslen-Wilson, 1985; Mitterer & Müsseler, 2013)] leading Mitterer & Müsseler (2013) to conclude that there would be "evidence for a loose perception–action coupling in speech" (p. 557).

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

These results are important because they are not compatible either with the hypothesis of a pure gestural or motor encoding of speech perception, or with a simple automatic sensory-motor coupling process. Indeed, if sensory-motor coupling were automatic, for translating sounds into gestures in the first case, or just to produce a rapid oral response in the second one, there would be no difference in response accuracy between oral and manual responses.

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

Whatever the interpretation, the data of Scarbel et al. (2014) show that close-shadowing gives access to valuable information as to the speed and efficiency of the internal sensory-motor relationship in the human brain. The present study is specifically devoted to the use of this paradigm to study perceptuo-motor interactions in distinct populations, for some of which perceptuo-motor linkage should be altered due to sensory degradation. The first and principal group of interest of this study is constituted by post-lingually deaf or hearing-impaired adults equipped with a cochlear implant (CI). For post-lingually deaf subjects, the perceptuo-motor linkage, acquired during speech acquisition before deafness, is still efficient during deafness as evidenced by the ability to maintain intelligible speech (Perkell et al. 2000). After CI, the sensory-motor coupling has to deal with a new auditory input provided by the CI. The major question addressed in this study consists of exploring through close-shadowing how sensory-motor relationships are reorganized after CI, comparing prior and present sensorial representation received by the patient.

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

Perceptuo-motor interactions in Cochlear-Implanted subjects

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

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

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

Perceptuo-motor interactions in normal-hearing elderly subjects

Since post-lingually deaf CI patients are often senior adults, we considered it to be relevant to compare them with an NHE population. The auditory abilities of elderly people decline as aging progresses (Gordon-Salant, 1986; Strouse at al., 1998). Specifically, even with close-to-normal auditory thresholds, normal-hearing elderly people have difficulties in adverse listening conditions such as in a noisy environment (Gelfand et al. 1985; Ohde & Abou-Khalil 2001; Fullgrabe et al. 2013). However, it is still unclear how cognitive decline might influence speech perception. Some studies showed a rather small influence of age-related decline on speech perception in noise (Cienkovski 2002; Cienkovski & Vasil 2010), although a general trend seems to exist of correlation between working memory and speech intelligibility in noise for elderly people (see a review in Füllgrabe & Rosen, 2016)

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

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

Objectives of the present study

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

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

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

MATERIALS AND METHODS

Participants

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

Insert Table 1 about here

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

Stimuli

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

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

The audio tracks of the stimuli were sampled at 44.1 kHz and presented without noise for CI, NHE and NHY participants. For NHY-noise participants, the 12 stimuli were mixed with white noise, low pass filtered at -6dB /oct, with a signal to noise ratio at -3dB (the signal energy being defined from burst onset to the end of the vowel). In the audiovisual modality of the experiment, the video stream consisted of 572-by-520 pixel/images presented at a 50 Hz rate with the speaker's full face presented with blue lips to exaggerate movements of the labial dynamics (Lallouache 1990).

Experimental Procedure

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

Please insert Figure 1 here

All groups performed the experiment in a soundproofed room. Participants sat in front of a computer monitor at a distance of approximately 50 cm. The acoustic stimuli were presented at a comfortable sound level, the same one for all participants (around 60 dB Sound Pressure Level). For CI, NHE and NHY participants, the presentation of acoustic stimuli was using a loudspeaker, whereas the presentation of acoustic stimuli was with earphones for NHY-noise participants. This procedure was required because noisy stimuli made acoustic analyses complex and possibly inaccurate if the stimulus and oral feedback were mixed. The presentation software (Neurobehavioral Systems, Albany, CA) was used to control stimulus presentation and to record key answers in the manual task. All participants’ productions were recorded using an AKG 1000S microphone for off-line analyses, with a system that ensured synchrony between the stimulus presented to the participant and the participant’s response. A brief training session preceded each task. The total duration of the experiment was around 30 min.

Acoustic Analyses

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

RTs were determined from the burst onset of the stop consonant to categorize the onset of the response (see Figure 1). In the manual decision task, the onset of the response corresponded to the key press. In the close-shadowing task, the burst onset of the stop consonant uttered by the participant in response to the stimulus was determined by inspecting the acoustic waveform and spectrogram information. RTs (reaction times) were only computed for correct answers: omissions or any type of errors (replacing one consonant by another or producing two consonants or two syllables in the close-shadowing task) were removed from the analyses.

Further response analyses

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

Then, to assess the relationship between accuracy and reaction time, an analysis of correlation between reaction time and percentage of correct answers was carried out for each group and each condition separately (oral and manual responses for audio and audiovisual stimuli).

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

RESULTS

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

Insert Figures 2 & 3 here

Normal-hearing elderly participants

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

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

Cochlear implanted participants

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

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

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

Comparison between the four groups

Reaction times (see Figure 2)

The main effect of group was significant ($F(3,45)=24.43$; $p<.001$), with slower reaction times for CI participants than for the three normal-hearing groups. No significant difference was observed between elderly participants and the two groups of young participants. However, NHY participants were faster than NHY-noise participants (overall, on average, CI: 648 ms; NHY-noise: 484 ms; NHE: 418 ms; NHY: 351 ms). The main effect of task ($F(1,45)=263.95$; $p<.001$) and of modality ($F(1,45)=21.09$; $p<.001$) were also found to be significant. For the task, oral responses were faster than manual ones (on average, 313 ms vs. 600 ms). For the modality, RTs were faster in the audiovisual modality than in the auditory modality (on average, 446 ms vs. 467 ms).

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

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

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

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

Percentage of correct responses (see Figure 3)

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

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

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

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

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

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

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

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

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

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

In the same figure, we superimposed data for the NHY group without noise onto that of the NHE group. It seems likely that the two groups behave differently, younger people being globally more accurate and faster than their elders (mean values for NHE: 89% and 418 ms; mean values for NHY: 95% and 351 ms). However, this was not displayed in the ANOVA, where no significant differences appeared between the two groups for either measurement, probably because of too small a number of subjects.

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

Insert Figure 4 here

Individual analysis in the CI group (see Table 2)

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

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

Insert Table 2 here

DISCUSSION

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

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

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

Effect of task: manual vs. oral responses

Normal-hearing elderly participants

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

Cochlear-implanted participants

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

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

Effect of the modality

Normal-hearing elderly participants

Regarding modality, it is important to notice that elderly participants presented response patterns that were quite similar to those of young participants without noise. Elderly participants did not present a benefit of the visual modality, since no difference in reaction times and in the percentage of correct answers were observed between responses for auditory and audiovisual stimuli. This lack of visual modality benefit might appear surprising, considering the relatively low accuracy scores in the auditory condition (around 80% globally in the oral response mode) since most studies on audiovisual perception in elderly people conclude that audiovisual integration abilities are quite similar in elderly participants as in younger ones (Cienkowski et al. 2002; Sommers et al. 2005, Stevenson et al. 2015), if not better (Sekiyama et al. 2014). This might suggest that audiovisual integration is less efficient in elder subjects under time pressure.

Cochlear-implanted participants

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

Interpretation in the framework of an analysis-by-synthesis process

Extending the discussion in our previous study (Scarbel et al., 2014), here we propose a tentative explanation of the difference between oral and manual responses, based on the neurobiological model of speech perception from Skipper et al. (2007), inspired from the analysis-by-synthesis approach (Stevens and Halle 1967) involving a perceptuo-motor processing loop between visual, auditory and motor areas in the human brain (Figure 5A). In this model, after a preliminary stage of unisensory processing respectively in visual and auditory areas, the auditory and visual information would converge in a multisensory area located in the posterior superior temporal cortex (STp), leading to a first multisensory phonemic hypothesis (stage 1 in Figure 5A). Articulatory goals and then associated motor commands corresponding to this initial prediction would be generated in frontal motor areas (stages 2-3), leading to the production of motor-to-sensory predictions (i.e. efference copy) and sent back to the temporal auditory cortex to be compared with the sensory auditory input (stage 4).

We hypothesized that oral and manual responses were generated at two different stages in this sensory-motor processing loop. Oral responses would be generated at an early stage, when motor commands are stimulated. This would make the oral answer faster, but also less accurate. Manual responses would be generated at the final stage of the perceptuo-motor loop. As a consequence, reaction times for manual responses would be slower than for the oral response, but the responses would be more accurate because of refined sensory-motor predictions.

Insert Figure 5 here

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

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

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

These results allow us to shed light on the perceptuo-motor linkage restoration in cochlear-implanted participants with a close-shadowing paradigm, resulting in faster but possibly less accurate responses in the speech task than in the manual task. Indeed, whatever the interpretation for the speed advantage in the oral task, it shows that cochlear-implanted subjects have recovered an efficient auditory-to-articulatory or motor connection. This connection enables them to efficiently convert the auditory input provided by the implant into the appropriate motor command. Strikingly, this ability is displayed soon after implantation, even after less than 3 months. Crucially, this restoration seems to be efficient in spite of the incomplete recovery of auditory perception abilities, since CI patients presented lower scores than normal-hearing participants.

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

Conclusion

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

Liberman, 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 Mod*elling (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.

Sumby, 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 Psychologia, 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.

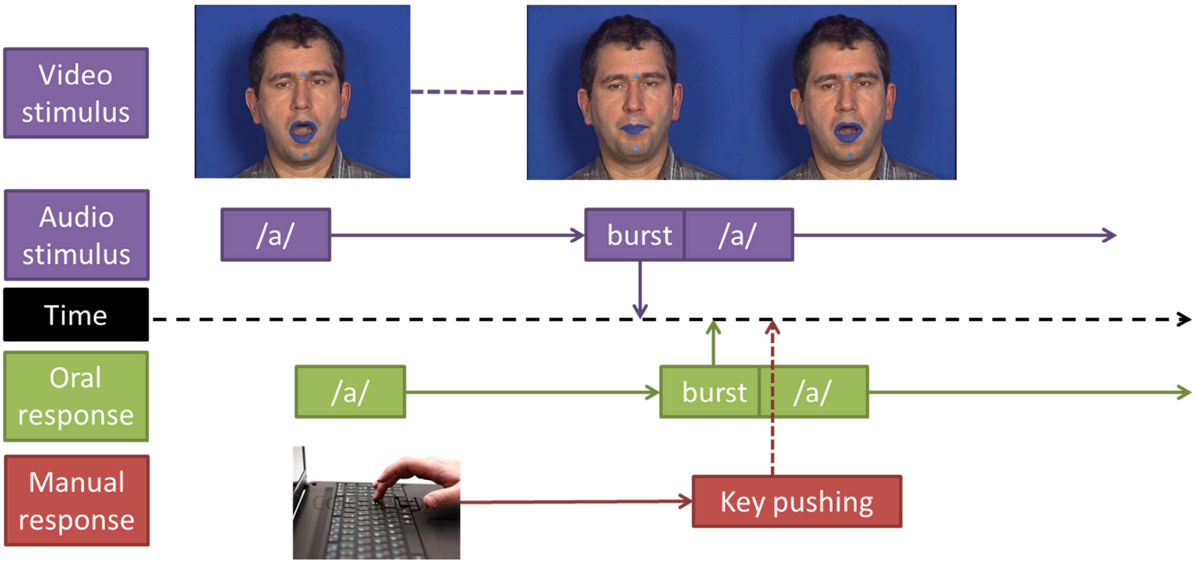


Figure 2: Mean RTs (in ms) in each group (normal-hearing young participants with noise (NHY-noise), normal-hearing young participants without noise (NHY), normal-hearing elderly participants (NHE), cochlear-implanted-participants (CI)), task (oral, manual response) and modality of presentation (auditory,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

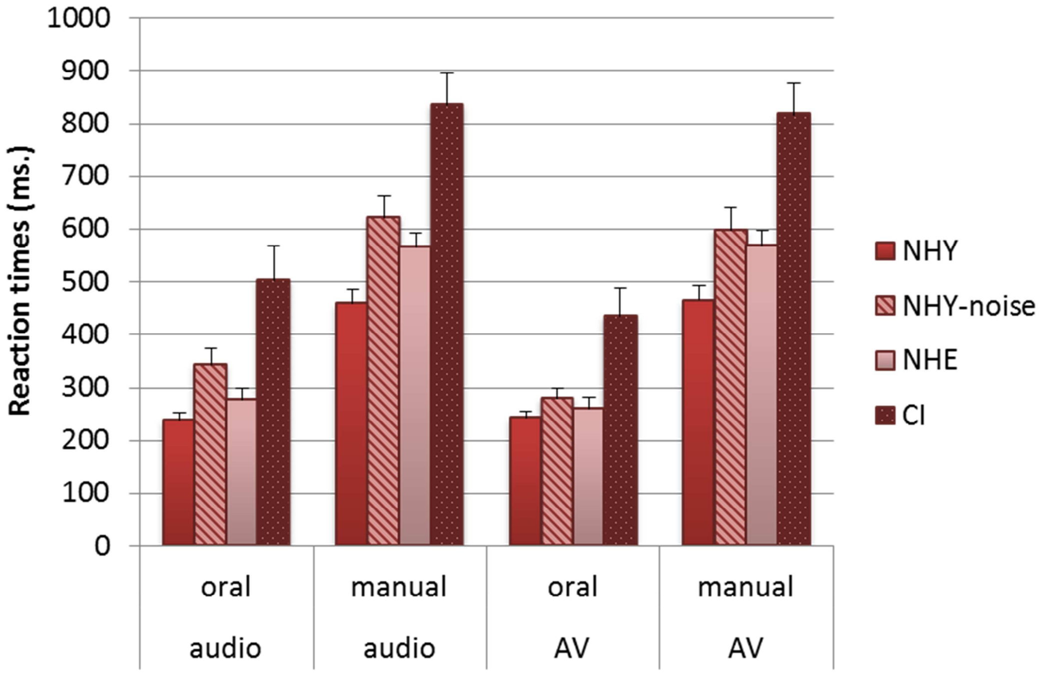


Figure 3: Mean percentage of correct identification in each group (normal-hearing young participants with noise (NHY-noise), normal-hearing young participants without noise (NHY), normal-hearing elderly participants (NHE), cochlear-implanted participants (CI)), task (oral, manual response) and modality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

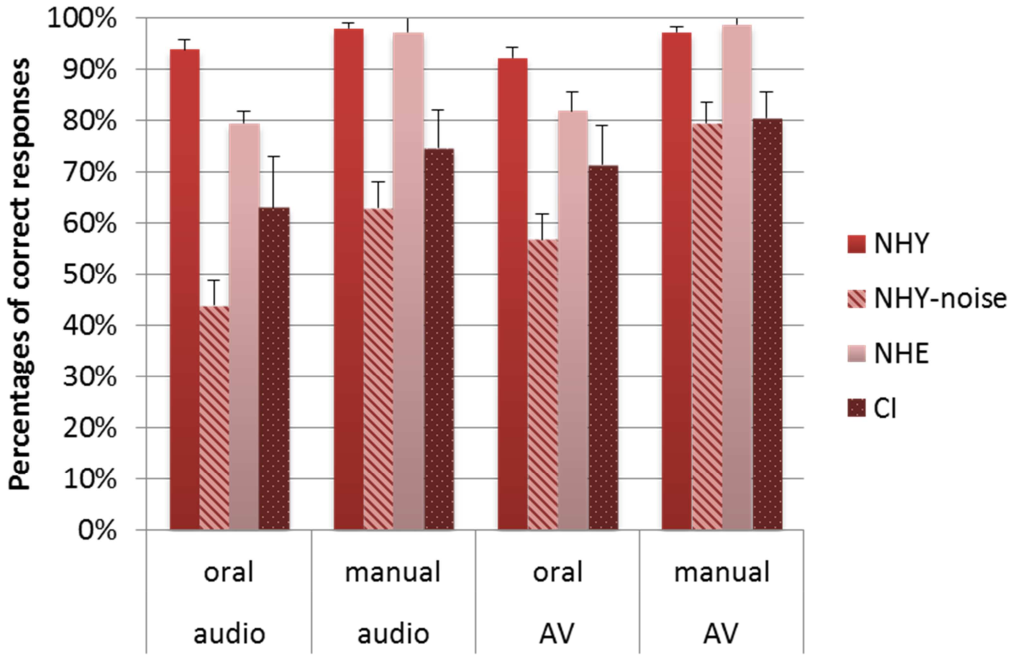


Figure 4: Correlation between RT (in ms) and percentage of correct answers (in %) in the NHE and NHY (A & B) and CI (C & D) groups, in response to audio (A & C) and audiovisual (B & D) stimuli. Correlation was significant for NHE subjects in the oral AV condition, with a positive slope. It was significant 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.

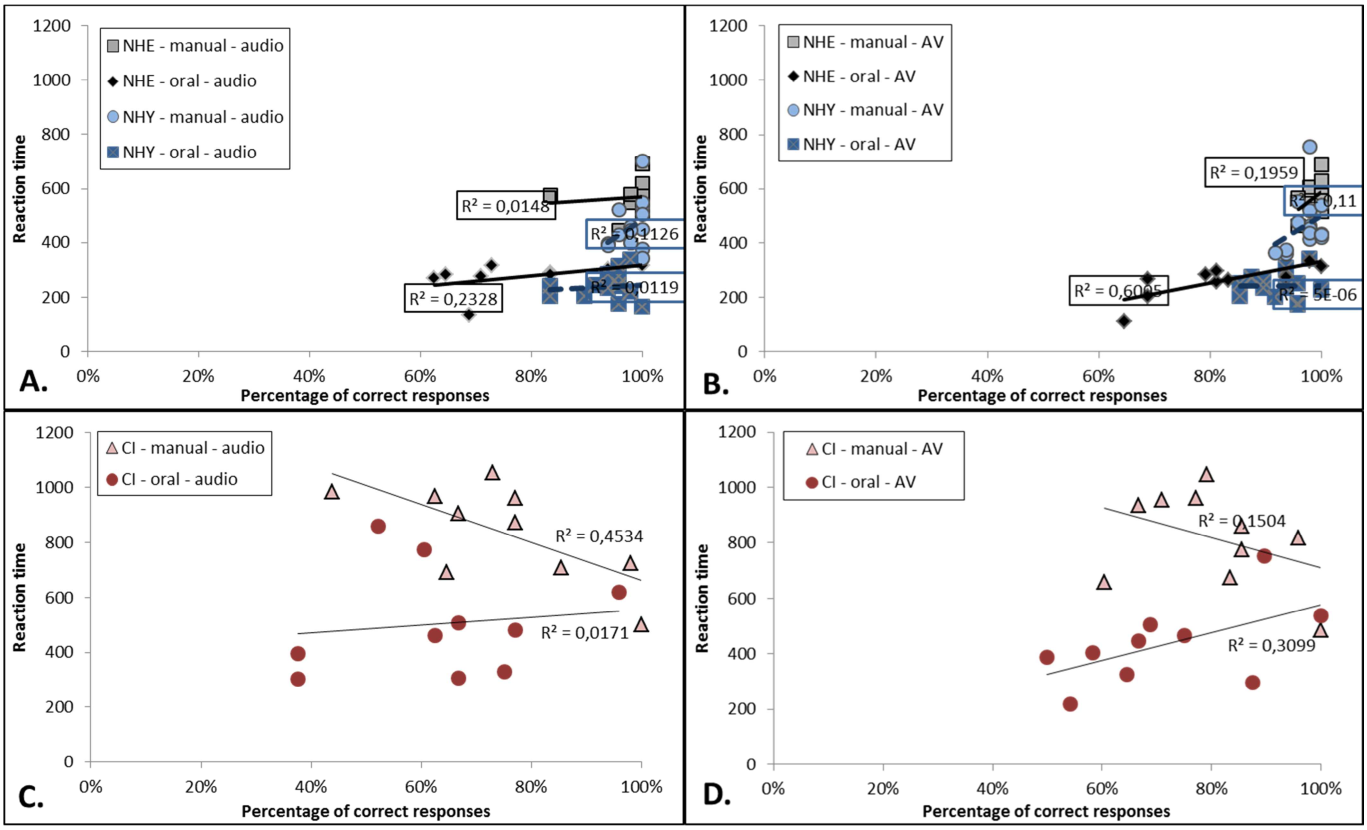
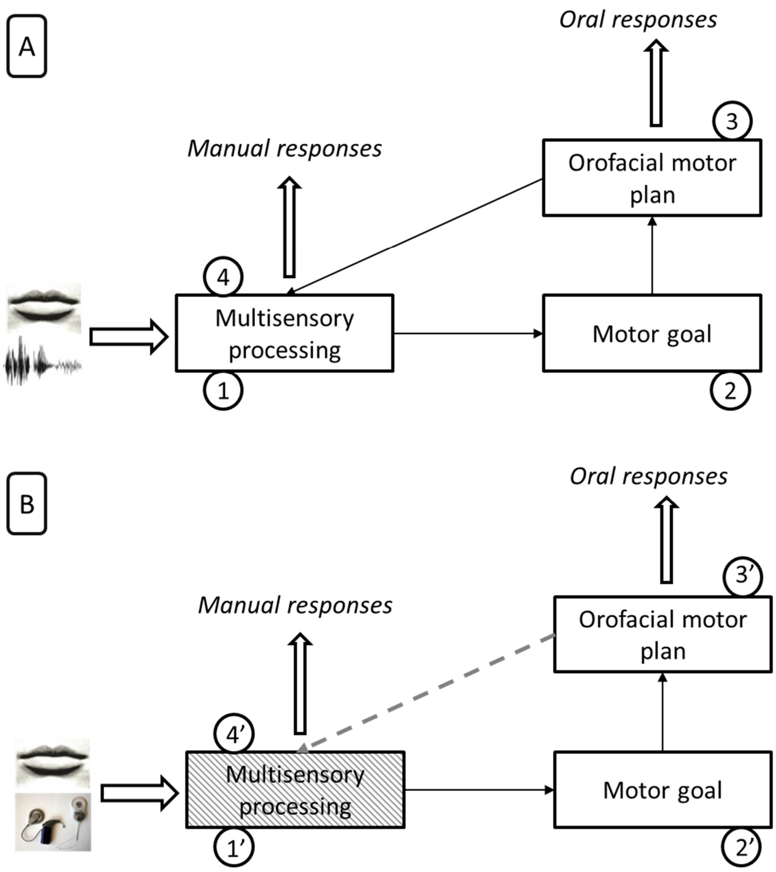


Figure 5: Possible auditory-motor relationships according to manual and oral response modes for normal-hearing participants (A) and cochlear-implanted participants (B) (adapted from Skipper et al.'s model, 2007). Multisensory processing would be degraded in CI patients (hatched on the figure).



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.