Cued Speech Enhances Speech-in-Noise Perception

Clémence Bayard, Laura Machart, Antje Strauss, Silvain Gerber, Vincent Aubanel, Jean-Luc Schwartz

Journal of Deaf Studies and Deaf Education, 2019, 1–11

doi:10.1093/deafed/enz003

Empirical Manuscript

Cued Speech Enhances Speech-in-Noise Perception

Clémence Bayard[1], Laura Machart[1], Antje Strauß[2], Silvain Gerber[1], Vincent Aubanel[1], and Jean-Luc Schwartz*[1]

[1]GIPSA-lab, Univ. Grenoble Alpes, CNRS, Grenoble INP

[2]Zukunftskolleg, FB Sprachwissenschaft, University of Konstanz

*Correspondence should be sent to Jean-Luc Schwartz, GIPSA-Lab, 11 rue des Mathématiques, Grenoble Campus, BP46, 38402 Saint Martin D'Hères Cedex, France (e-mail: jean-luc.schwartz@gipsa-lab.grenoble-inp.fr)

Abstract

Speech perception in noise remains challenging for Deaf/Hard of Hearing people (D/HH), even fitted with hearing aids or cochlear implants. The perception of sentences in noise by 20 implanted or aided D/HH subjects mastering Cued Speech (CS), a system of hand gestures complementing lip movements, was compared with the perception of 15 typically hearing (TH) controls in three conditions: audio only, audiovisual and audiovisual + CS. Similar audiovisual scores were obtained for signal-to-noise ratios (SNRs) 11dB higher in D/HH participants compared with TH ones. Adding CS information enabled D/HH participants to reach a mean score of 83% in the audiovisual + CS condition at a mean SNR of 0 dB, similar to the usual audio score for TH participants at this SNR. This confirms that the combination of lip-reading and Cued Speech system remains extremely important for persons with hearing loss, particularly in adverse hearing conditions.

Auditory speech perception for deaf or hard-of-hearing persons

In recent years, a large number of deaf people (and in particular many congenitally deaf children) are fitted with a hearing aid (HA) or a cochlear implant (CI). With technological progress, CIs have become the most effective vehicle for helping profoundly deaf people to understand spoken language, to perceive environmental sounds, and, to some extent, to listen to music. The development of the early detection of deafness together with the trend for more and earlier implantation might decrease or minimize the interest for visual cues in oral communication. HAs and CIs immensely help deaf people by providing auditory access to speech. Yet, the auditory input they deliver remains degraded compared to the full auditory signal (Shannon, Fu, Galvin, & Friesen, 2004; Percy et al., 2013; Wolfe, Morais, Schafer, Agrawal, & Koch, 2015; Todorov & Galvin, 2018).

This is particularly the case for speech perception in noise, which remains a really difficult task for deaf people (Revoile, Pickett, & Kozma-Spyteck, 1991; Zeng & Galvin, 1999; Caldwell & Nittrouer, 2013; Srinivasana, Padilla, Shannon, & Landsberge., 2013). HAs and CIs provide inaccurate representations of phonemically relevant spectral structure making the perceptual segregation of that spectral structure from background noise difficult (Baer, Moore, & Gatehouse, 1993; Boothroyd, Mulhearn, Gong, & Ostroff, 1996; Fu, Shannon, & Wang, 1998; Bernstein & Brungart, 2011). Friesen, Shannon, Baskent, & Wang (2001) showed that CI users who displayed incremental benefit beyond 4 channels tended to do better in noise than CI users who only showed growth with increase up to 4 channels. As a consequence, they estimate that CI users require a minimum of 8-10 independent spectral channels to perceive speech in noise. In comparison, understanding speech in quiet requires only 4 spectral channels (Shannon, Fu, & Galvin, 2004). Since CIs have typically between 12 and 22 physical electrodes, this might appear sufficient for processing speech in noise, but it appears that CI listeners perform as if information were provided by only 4 to 8 independent information channels

(Friesen et al. 2001; Strauß, Kotz, & Obleser, 2013). The lack of fine temporal structure also contributes to the difficulty to segregate a target signal from background noise (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006).

The role and potential limitations of lipreading

Understanding speech in noise is of course crucial considering that noisy environments are more likely to occur than clean ones in most situations in real life. Therefore, visual information provided by lip-reading remains paramount for Deaf/Hard of Hearing people (D/HH in the following), be they fitted with HA or CI. Starting with pioneering studies done by Norman Erber or others in the 1960s to early 1970's (Erber 1975), it is now well known that D/HH listeners spontaneously process lip-reading information to compensate for their auditory deficit (Lachs, Pisoni, & Kirk, 2001; Bergeson, Pisoni, & Davis, 2005). As a matter of fact, in spite of a large inter-individual variability, D/HH persons happen to be among the best lip-readers (Bernstein, Demorest, & Tucker, 1998, 2000). They can display significant audiovisual fusion after cochlear implantation (Schorr, Fox, van Wassenhove, & Knudsen, 2005; Rouger et al., 2007) and in some cases may even be better audiovisual integrators than individuals with typical hearing (Rouger et al., 2007).

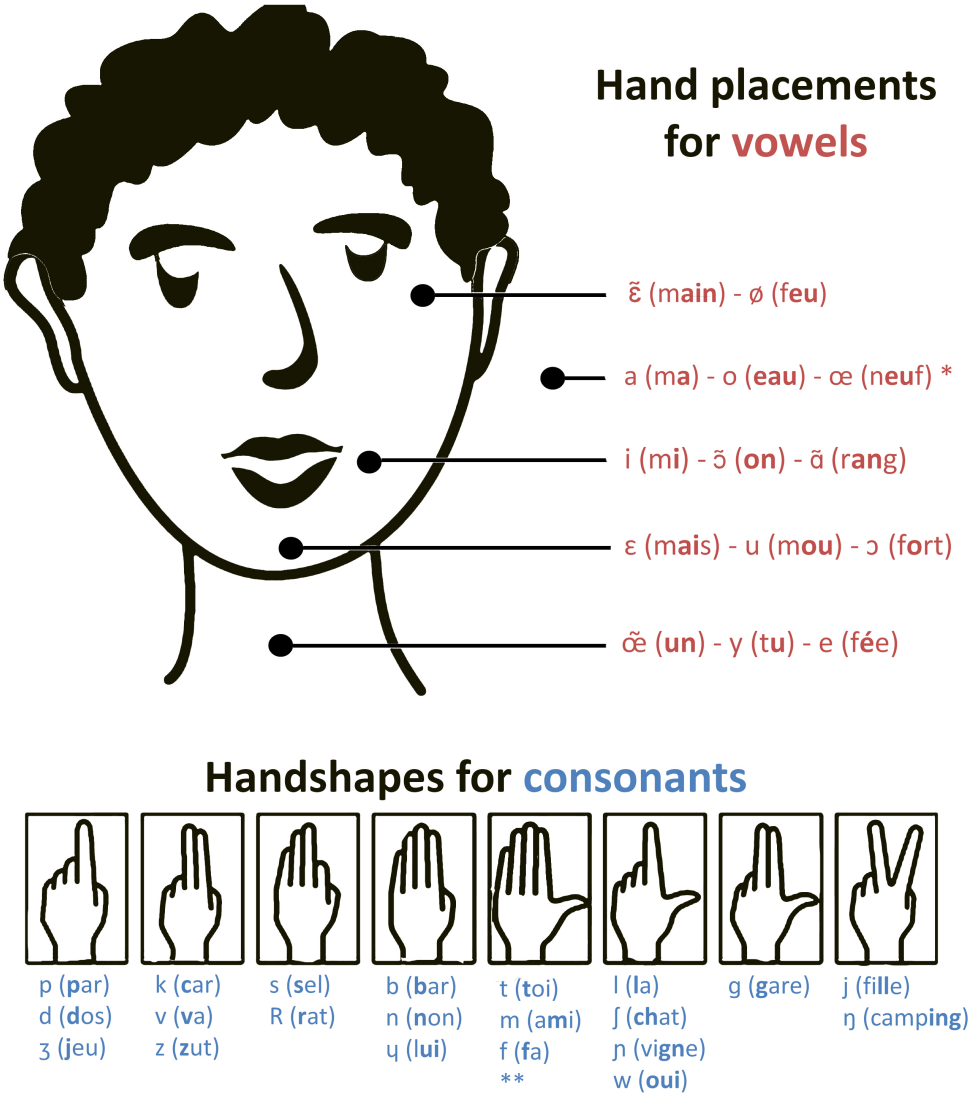
Still, while speech-reading can help mitigate missing auditory information to some extent, speech-reading information alone is likely insufficiently rich to overcome the severe degradation posed by noise to CI deaf participants. Surprisingly, this has actually seldom been tested. As a matter of fact, a number of studies have been done on audiovisual speech perception in noise in unaided D/HH participants (e.g. Erber 1971; Grant, Walden, & Seitz,1998; Bernstein & Grant 2009); or on auditory-only speech perception in noise in aided or cochlear-implanted D/HH participants (e.g. Caldwell & Nittrouer 2013; Tabanez do Nascimento & Bevilacqua 2005); or on audiovisual speech perception

without noise in aided or cochlear-implanted D/HH participants (e.g. Holt, Kirk, & Hay-McCutcheon, 2011; Liu et al., 2014). But there are only a small number of studies reporting audiovisual speech perception in noise in aided or cochlear-implanted D/HH participants. Leybaert & LaSasso (2010) report inaccurate audiovisual perception of vowel-consonant-vowel (VCV) sequences embedded in acoustic noise by deaf CI children. Taitelbaum-Swead & Fostick (2017) obtain lower speech perception accuracy for monosyllabic words in white noise at a signal-to-noise ratio (SNR) of 0 dB for CI children and adults as compared to typically hearing participants of the same age, for both auditory and audiovisual presentations.

Cued Speech, a potential crucial complement for speech communication in noise

In this context, it is essential to provide D/HH persons with additional support for communication, particularly in adverse conditions. It has been proposed 50 years ago that the addition of manual cues from the Cued Speech (CS) system could help D/HH individuals to overcome the uncertainty of auditory signals delivered by the CI or HA. Originally, this system was designed to help deaf individuals (without CI) to perceive speech through disambiguating the visual modality (Cornett 1967). The CS system resolves the ambiguity in lip-reading by making each of the phonological contrasts of oral language visible. Each syllable is uttered with a complementary gesture called a manual cue. In its French version, vowels are coded with five different hand placements in relation to the face, and consonants are coded with eight handshapes (see Figure 1). Each manual cue can code several phonemes, but these phonemes differ in their labial visual pattern. Moreover, consonants and vowels sharing the same labial visual pattern are coded by different cues. Therefore, both sources of information (hand and lips) provide complementary information. Cued Speech has been shown to enhance the development of speech perception and language processing in CI children (Leybaert &

100 LaSasso 2010) and to constitute, in addition to audition and lip-reading, a common amodal network
101 for language processing in the brain (Aparicio et al. 2017).



102

103 **Figure 1.** Cues in French Cued Speech: hand shapes for consonants and hand placements for vowels.

104

When communicating with CS, a person talks while simultaneously cueing. The combination of visual information, provided by the articulatory labial movements and manual cues, allows deaf individuals to perceive all syllables efficiently (Clark and Ling 1976; Nicholls and Ling 1982; Gregory 1987; Périer, Charlier, Hage, & Alegria, 1990; Uchanski et al., 1994; Bratakos, Duchnowski, & Braida, 1998; Alegria & Lechat, 2005). For example, Nicholls & Ling (1982) tested the identification of syllables and keywords within sentences, by 18 deaf children from 9 to 17, with at least 4 years of experience with CS. Syllables comprised all combinations of CV and VC syllables with C one of the 24 consonants and V one of the three /i a u/ vowels in American English. They obtained syllable recognition scores as high as 80% when CS was available, and less than 40% when it was not. Speech reception was higher than 95% for keywords.

Moreover, the CS system enables the perceiver to focus attention in time. Indeed, whereas Cornett (1994) described the CS system as a time-locked system characterized by a synchrony between manual cues and sounds, Attina, Beautemps, Cathiard, & Odisio (2004,) and Attina, Cathiard, & Beautemps (2006) found that the sounds and hands were not synchronous, as it is also the case for sounds and lips in audiovisual speech (Schwartz & Savariaux 2014). Manual cues naturally precede sound, since the hand reaches its target position and shape up before the vowel target in consonant-vowel syllables (with an advance estimated on French CS speakers to 200 ms by Attina et al., 2004 or Gibert, Bailly, Beautemps, Elisei, & Brun, 2005; and with a smaller advance of 100 ms used by Duchnowski et al., 2000, in their automatic system for American CS synthesis). Interestingly there is still a rather precise temporal coordination between speech and manual cues, but it is in advance of the sound. Indeed, Attina et al. (2004) observed that the hand reaches its position for a given consonant-vowel syllable precisely at the temporal position of the consonant constriction. Deaf individuals were shown to take

advantage of the advance of manual cues relative to lip-reading cues during CS perception (Attina, 2005; Troille, Cathiard, & Abry, 2007).

Strikingly, most studies assessing the role of CS were realized in a pure visual environment without sound. The combination of sound, lips and manual cues was only recently explored by Bayard, Colin, & Leybaert (2014) and Bayard, Leybaert, & Colin (2015) who examined syllable perception by CI participants in a paradigm including various cases of congruent or incongruent combinations of auditory and visual speech stimuli. The results showed that, in quiet conditions, CS receivers do combine sound, lip shapes and manual cues into a unitary percept. Still, no study attempted, to our knowledge, to assess the potential benefit provided by CS to improve speech perception in noisy conditions. A D/HH person communicating in a noisy environment and receiving CS information from a partner mastering this system has to solve a complex processing-and-fusion problem. Indeed, adequate reception involves (i) efficient processing of the auditory input degraded by noise, exploiting the benefit of the cochlear implant or hearing aid, and (2) fusion of three sources of information that are sound, lips and hands. Yet it is not known at this stage how D/HH participants can deal with this complex task, the more so in environments where the structure of information may change from time to time (e.g. communicating with partners who either master and use CS or who don't, hence possibly switching between different kinds of fusion situations in the course of communication).

The objective of the present study is to evaluate this capacity in more detail. For this aim, we assessed the comprehension of sentences in noise by a group of D/HH CS users compared to a group of typically hearing (TH) controls, in three conditions: audio only, audiovisual and audiovisual with CS. It is well known that D/HH participants are very heterogeneous concerning their auditory abilities. For this reason, and to facilitate further comparison between groups, we customized the SNR for each

participant, ensuring that their level of correct comprehension was about 60% in the audiovisual condition.

With this study, we aimed to answer two basic questions. First, we wanted to evaluate whether there was indeed a difference in the SNRs enabling to achieve similar levels of sentence comprehension across D/HH and TH participants in the audiovisual condition. In fact, there are almost no data in the literature assessing the reception of audiovisual speech in noise in HA or CI D/HH persons with complete sentences, though this is actually a crucial task for assessing their comprehension capacities for speech communication. Second, we wanted to investigate whether CS does provide a gain in the perception of audiovisual speech in noise for D/HH CS users, just as it does when there is no auditory input at all. Particularly, we wanted to check their ability to efficiently integrate the three sources of information (noisy sound, lips and hands), in a paradigm mixing conditions in a single block, imposing the participants to permanently monitor their attention and modulate the fusion process accordingly.

Material and Methods

Participants

The recruitment of D/HH participants fitted with a cochlear implant or a hearing aid and mastering CS is rather complicated and slow. The recruitment of D/HH participants mastering CS in this study was considerably facilitated by the opportunity provided by a week of CS training organized by the French ALPC association. This association supports the use of the French CS version for communication between hearing persons and persons with an auditory handicap (http://alpc.asso.fr/). Twenty D/HH teenagers and adults (nine female; age range 12–21 years, mean = 15.5; see Table 1 for more details) participated in the study. Nineteen participants had a profound deafness and the remaining one a severe

171 deafness. Seventeen of them were cochlear implanted (six bilaterally), the remaining three being fitted
172 with HA.
173 All participants had been using CS receptively ("decoding") on average since 4.5 years of age, and
174 they learned to cue ("cueing") on average since the age of 5.5 years. The information was provided by
175 the participants, and hence corresponded to a rough self-estimation of these behavioural abilities.
176 Notice that while in most cases decoding was used before coding, the order could be different in rare
177 cases, particularly for participant 14, probably because this participant had a deaf brother, and a hearing
178 impairment which increased with age in his first years of age. They all communicated with their
179 environment orally, being able to both understand speech from the sound and sight of their interlocutor,
180 and pronounce intelligible speech that their interlocutor could understand. They were integrated in a
181 family and school environment mostly comprised of typically hearing family members or school
182 colleagues with whom they communicated orally without CS. They frequently used CS for decoding
183 language (understanding), typically with parents or school assistants providing on-line CS in school,
184 and particularly in noisy environments. Expressively, they cued while speaking much less frequently.
185 In one case, the participant cued expressively only in the course of the training sessions organized by
186 ALPC.
187 To evaluate the speech perception performance of this group of adolescent-to-young-adult D/HH
188 participants, we compared their performance with a reference group of TH adult participants. We could
189 have used as a control a TH group matched in age, but we preferred using TH adults with a completely
190 mature auditory/cognitive system, to provide an optimal baseline enabling to better evaluate the deficit
191 in perception for the group of D/HH participants in adverse conditions. Typically hearing participants
192 were recruited by announcements in the *** and on a national website
193 (http://expesciences.risc.cnrs.fr/pre_formulaire.php), and tested at ***. Fifteen self-reported TH adults

(nine female; age range 22–36 years, mean = 29.5; see Table 2 for more details) participated in the study.

Written informed consent was obtained from each TH and D/HH participant together with parental authorization for minors. The experiment was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and was validated by the *** Ethics Board (***). All participants were French native speakers with normal or corrected-to-normal vision and did not have any declared language or cognitive disorder.

Insert Tables 1, 2 here

Stimuli

Sentence material was selected from the Fharvard corpus (Aubanel, Bayard, Strauß, & Schwartz, submitted), a French equivalent of the Harvard sentences (Rothauser et al., 1969), which have been used extensively in speech perception research (e.g. Bradlow, Torretta, & Pisoni, 1996; Cooke et al., 2013). The corpus consists of 70 phonemically-balanced lists of 10 sentences, where each sentence contains five keywords used for scoring. Each keyword contains one or two syllables and is relatively poorly predictable from context (e.g. "Elle *attend* le *taxi* sur la *pelouse devant* l'*hôtel*" [engl. "She is *waiting* for the *taxi* on the *grass* in *front* of the *hotel*"]; keywords in italics). For this experiment, a subset of 42 different sentences from the Fharvard corpus was presented.

The 42 sentences were spoken by a professional cuer (female, 34 years old, highly experienced in CS production, with a French diploma, "CS professional degree" followed by 8 years of professional practice, enabling her to assist teachers by CS production in a school for D/HH children). Each sentence was recorded two consecutive times: the speaker first produced the sentence with manual cues (AVC)

and then without cueing (AV). All productions were checked for accuracy by another professional CS cuer, and some productions were removed when a gesture was inaccurate or seemed ambiguous.

It is well known that there is a trend that speech rate is slower with CS than without CS (Attina et al. 2004). To be able to compare speech perception in noise with and without CS, we had to control for this difference. If production was slower, it might increase intelligibility in the CS condition. Therefore, we asked the cuer to maintain a stable production rate during recording as much as possible. Afterwards, we systematically evaluated the duration of recorded sentences, and selected 28 out of 42 sentences, in which both recordings with and without CS production had similar speech rates. The mean duration of selected sentences was 3597 ms for AVC and 3551 ms for AV (3.76 vs. 3.81 syllables/s). The mean difference of 45 ms between AVC and AV productions was not significant ($t = 1.65$, $p > .05$). The remaining 14 sentences (pronounced either with or without CS) were used to prepare the Audio-only condition by extracting audio files and dubbing them on a video of the cuer with a neutral face (see Figure 2). Video editing was realized with the Adobe Premiere software.

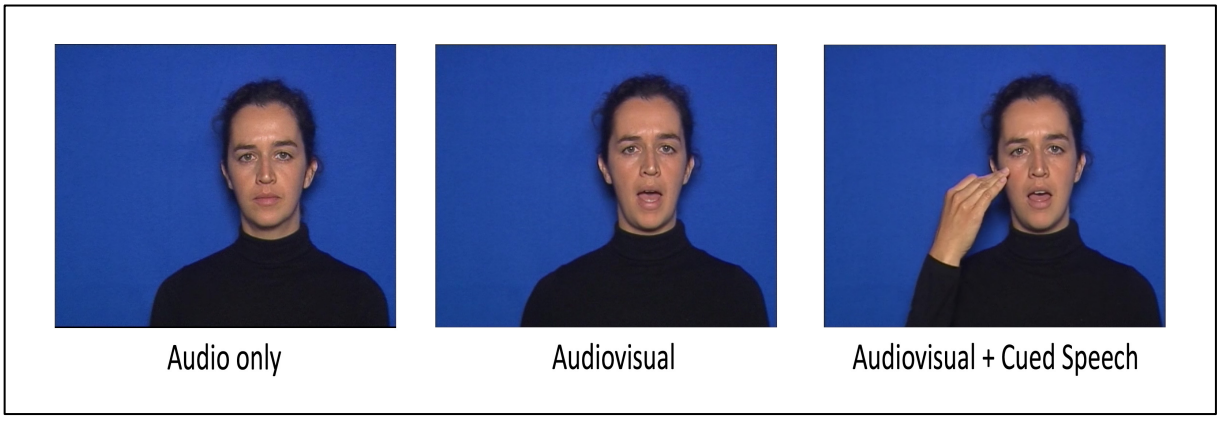


Figure 2. Illustration of the three experimental conditions: Audio only (A), Audiovisual (AV) and Audiovisual with Cued Speech (AVC).

Each participant was presented with all 42 sentences in noise: 14 Audio only (A), 14 Audiovisual (AV), and 14 Audiovisual with Cued Speech (AVC). To control for item effects, we created two orders counterbalancing sentences between conditions AV and AVC (see Table 3). Half of the participants were presented with sentences in order 1 and the other half with sentences in order 2. For example, sentences presented in the AV condition in order 1 were presented in the AVC condition in order 2, and vice-versa. In order 1, half of the sentences in the A condition contained extracted sounds from the AV production and the other half from the AVC production, and conversely in order 2. All 42 stimuli were presented randomly to each participant in one block, hence A, AV and AVC conditions were mixed in an unpredictable order.

Insert Table 3

To generate speech-shaped noise, white noise was filtered by the long-term average speech spectrum taken from the whole recorded corpus. Each sentence was embedded in noise that started 500 ms before and ended 500 ms after the sentence. 24 different signal-to-noise-ratio values (SNR) from +6 dB to -17 dB in 1 dB steps were prepared to enable the selection of the adequate SNR for each listener in a pre-test procedure (see "Procedure").

Procedure

The experiment was conducted in a quiet room with the Presentation ® software (www.neurobs.com). Instructions and videos were displayed on a laptop at eye level and approximately 70 cm from the

participant's head. Sound was played with loudspeakers (Logitech) at a comfortable level around 70 dB SPL (similar for all participants). The D/HH participants used the standard setting of their cochlear implant or hearing aid, with no quantitative evaluation of this setting.

Participants completed pre-testing prior to the main test battery. In the pre-test, the SNR for the main experiment was determined for each participant individually. Mimicking an adaptive tracking procedure, groups of three sentences not used in the main experiment were presented first at a high SNR in the AV condition and participants had to repeat what they had heard. The experimenter calculated the percentage of correct responses and decreased the SNR by 1 dB step for the next sentence group until the number of correctly recognized words fell below 10. A recognition score of 60 % corresponds to 9 out of 15 correctly recognized keywords. The final SNR value for each participant was selected when the participant correctly recognized between 9 and 10 words twice.

In the main experiment, participants were asked to listen to and look at the video and, after each sentence, to repeat aloud what they had heard. The experimenter wrote on a sheet the number of keywords correctly recognized from 0 to 5. The experimenter was facing the participant and could not see the screen. Hence, since all conditions were randomized in a single block, the experimenter did not know the condition being tested while scoring the corresponding comprehension score. The experimenters were trained to the possible speech disfluencies in the production of the D/HH participants, hence the scoring was straightforward for these participants as well as for the TH ones. The total duration of the experiment, pre-test and main experiment included, was approximately 25 minutes.

283 **Statistical analyses**

284 A first question asked in this study concerned differences in SNR values in the pre-test for TH vs.
285 D/HH participants. The assumption was that SNR values associated to AV speech reception around
286 60% would be lower for TH than for D/HH participants. Because of the large inter-individual
287 variability in the D/HH group, classical in all studies assessing comprehension performances, it appears
288 that these values were not distributed according to a Gaussian law. Therefore, the difference in SNR
289 values between the two groups was evaluated by a Wilcoxon sign-rank test.

290 A second set of question concerned differences in speech reception between conditions (A, AV, AVC)
291 and between groups (TH, D/HH). We expected lower reception scores in the A than in the AV condition
292 for both groups, and higher scores in the AVC condition for the D/HH group, and possibly also, to a
293 lesser extent, for the TH group. Indeed, considering that TH participants have no experience with Cued
294 Speech, there might appear in the experiment some learning processes according to which the reception
295 scores would increase along the experiment for TH participants in the AVC condition. Moreover,
296 learning effects could also appear for both groups and all conditions just in relation with task learning.
297 Hence, in addition to the variables GROUP (TH, D/HH) and CONDITION (A, AV, AVC) we added
298 a variable TRIAL with 14 values corresponding to the number of the trial (from 1 to 14) in the test in
299 a given condition. The dependent variable was the number of correct keywords from 0 to 5, considered
300 as a categorical ordered variable with 6 levels. Participants were considered as a random factor, in a
301 mixed design with GROUP as a between-subject factor and CONDITION and TRIAL as within-
302 subject factors. The effects of these four factors (PARTICIPANT, GROUP, CONDITION, TRIAL)
303 were assessed by an ordinal regression with random effects (Tutz & Hennevogl, 1996), by using the
304 *clmm* ordinal package in the R (version 3.2.0) software (R Development Core Team, 2016).

Selection of the appropriate model was based on log-likelihood differences between models, assessed with a Chi-square test with a degree of freedom equal to the difference in the number of parameters, and with the criterion of p-value lower than 0.05. The analysis of reception scores was done in two steps, first for selecting random effects and then for selecting fixed effects. In the first step we tested the need for a PARTICIPANT-CONDITION or a PARTIPANT-GROUP random effect, assessing whether individual variability differed when the participants moved from one modality to another of the CONDITION factor or between TH and D/HH participants. Then we studied the structure of fixed effects by a descendant analysis with the *anova* function in R. At the end of this process, we checked by graphical inspection of residuals if the model adjusted the data correctly, i.e., if the empirical probabilities were close to the probabilities estimated by the model and if they were within the range of 95% prediction. Finally, multiple comparisons were achieved taking into account that we use an ordinal regression model with random effects. They were realized using the *lsmeans* function of the *lsmeans* package of the R software. This method ensures that the risk of type I error does not exceed 0.05.

Results

Differences in AV perception between the two groups

On Fig. 3, we present the individual values of SNR and AV scores for all D/HH and TH participants (see also individual values per participant in Tables 1 and 2). As expected, the mean AV score is around 60% for both populations, although the dispersion of scores is actually rather large. Indeed, performance may be rather variable from one sentence to another, and it appears that the initialization phase was insufficient for a perfect estimation of the SNR ensuring a 60% performance for each participant. Still, most participants have their performance in the 40-80% range (with only one TH participant and 3 D/HH participants having scores below 40%). Linear regression lines relating SNR

and AV score in each group (in dotted lines on Fig. 3) show that the mean AV score is rather stable, around 55%, along SNRs for both D/HH and TH participants.

The mean SNR for TH controls was –11.2 dB, while the mean SNR for D/HH participants was –0.1 dB (11 dB higher). In more detail, the range of SNRs for TH participants varied from -14 to -9 dB (Table 2). It varied for D/HH participants from -11 to 6 dB with much larger variability, and apart from one single participant with an SNR at -11 dB (that is within the range of values for the TH group) all SNR values were higher than -8 dB for the 19 remaining D/HH participants. The difference between SNR values in the TH vs. D/HH group is highly significant (Wilcoxon sign-rank test, $Q=292.5$, $p=2.10^{-6}$).

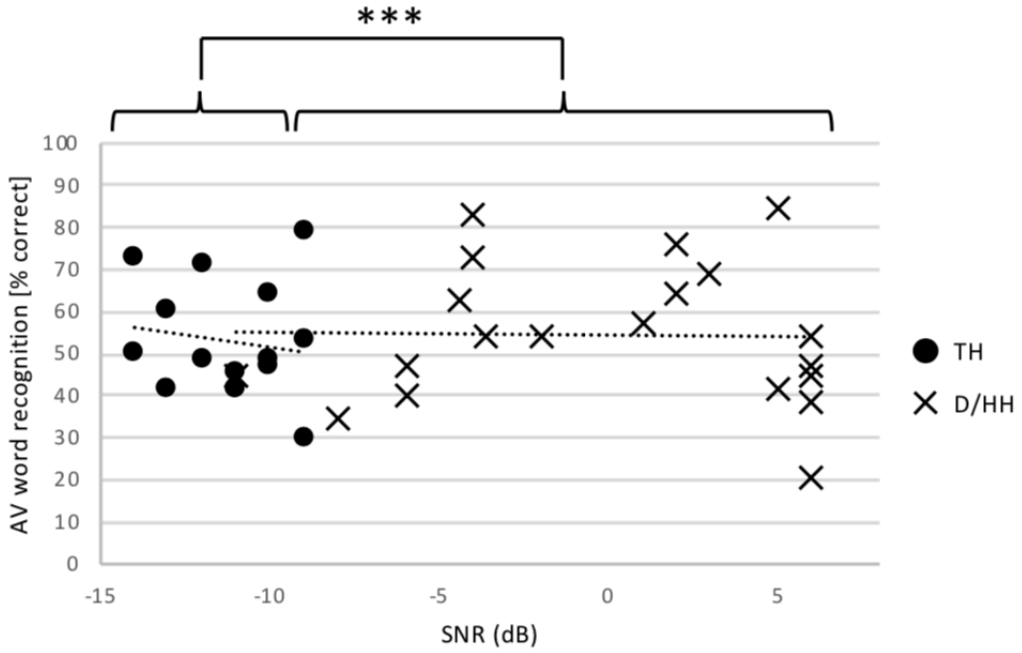


Figure 3. Individual values of selected SNR for a theoretical AV score at 60%, and actual AV score at the corresponding SNR, for each participant of both groups. *** refers to a highly significant difference ($p<0.001$). Dotted lines display the linear regression between SNR and AV score in each group.

342

343 **Differences in perception between conditions and groups**

344 For each participant and each condition, the percentage of correctly recognized words was calculated.
345 The mean scores for the two groups and the three conditions are displayed in Figure 4. They were
346 submitted to an ordinal regression test with three fixed factors GROUP, CONDITION, TRIAL and a
347 random factor PARTICIPANT according to the sequence of statistical analyses described previously.

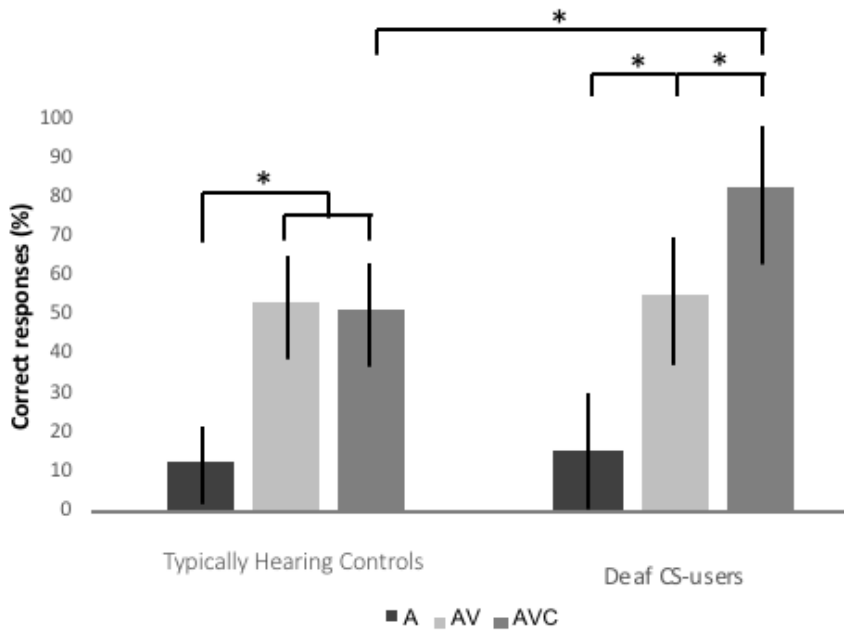
348 The analysis of random effects displays a significant PARTICIPANT-CONDITION interaction
349 showing that inter-individual variability differs in the three conditions. The corresponding values show
350 that variability is larger in the A condition, intermediate in the AVC condition and smaller in the AV
351 condition, with no significant difference in variability between groups.

352 The analysis of fixed factors hence incorporated this PARTICIPANT-CONDITION interaction. It
353 displayed no effect of TRIAL alone or in interaction. Hence there is no significant learning effect in
354 the results. The selected model included a highly significant GROUP-CONDITION interaction
355 ($\chi^2(2)=30.59$, $p<0.0001$).

356 The multiple comparison analysis confirmed that labial information significantly increased
357 comprehension for both D/HH and TH participants: scores in the A condition are lower than those in
358 the AV condition (15% vs. 55% for D/HH; z_ratio = 9.04; $p<0.0001$; 13% vs. 53% for TH; z_ratio =
359 8.06; $p<0.0001$). Furthermore, the difference between A and AVC was significant in both groups (TH:
360 z_ratio = 5.95; $p<0.0001$; D/HH: z_ratio = 11.34; $p<0.0001$). As expected, adding manual information
361 improved scores only for D/HH CS users, for which the performance increased from 55% in the AV
362 condition to 83% in the AVC condition (z_ratio = 9.23; $p<0.0001$). By contrast, the typically hearing

363 controls showed similar performance in the AV and the AVC condition (53% vs. 52%; z_ratio = 0.40;
364 p = 0.92).

365 The comparison between groups confirmed that both groups had similar scores in the AV conditions
366 (D/HH: 55% vs. TH: 53%; z_ratio = 0.20; p = 1), close to the targeted 60% value. The scores are also
367 similar across groups in the A condition (D/HH: 15% vs. TH: 13%; z_ratio = 0.08; p = 1), while the
368 scores in the AVC condition are indeed better for D/HH CS users than for TH participants (D/HH:
369 83% vs. TH: 52%; z_ratio = 5.33; p < 0.0001).



370

371 **Figure 4.** Percentage of correct responses by group and conditions. A = audio only, AV = audiovisual,
372 AVC = audiovisual + Cued Speech gestures. Stars display significant differences in the multiple
373 comparison analysis, with p<0.05 (see section about statistical analyses).

374

375 **Inter-individual variability in the benefit provided by the visual sources in the D/HH group**

In a last stage of analysis, we attempted to explore in more detail the correlates of differences between participants in the D/HH group, known for being highly inhomogeneous. To this aim, we used SNR as a proxy for characterizing the speech reception abilities of the D/HH participants, and assessed the relation of this proxy with performance in the three perceptual conditions. We observe in Figure 5 that there is actually an interesting portrait here. While perception scores in the AV condition are quite stable among SNR, which is not surprising since they were precisely controlled for being more or less constant around 60%, there is a trend that A perception increases and AVC perception decreases with increasing SNR. This suggests that the participants with most perception difficulties (highest SNR) reach a 60% score in the AV condition with already a rather high A score, hence benefit less of the visual modality for this performance. Moreover, they also seem to benefit less of the manual input in the AVC condition.

This is confirmed by a statistical test of the relation between SNR (considered to characterize the level of A recovery provided by the cochlear implant or hearing aid) and the gain provided by the two visual sources in the experiment (provided by lips and hands). Normalized gains were used in this analysis, that is the ratio (AVC – A)/(100 –A), with data in percentage, for evaluating the gain AVC vs. A, and the ratio (AVC – AV)/(100 –AV), with data in percentage, for evaluating the gain AVC vs. AV. Pearson correlations are significant for the gain AVC vs. A, evaluating the global benefit provided by the two sources ($r^2 = 0.223$, $t(18) = 2.269$, $p = 0.036$) and for the gain AVC vs. AV, evaluating the specific benefit provided by CS ($r^2 = 0.226$, $t(18) = 2.289$, $p = 0.034$).

It is important to notice that three participants in the D/HH group appear to reach rather low perception scores in the AVC condition, around 50% or even lower while it is higher that 70% for all other participants (see Fig. 5). This is not related to specificities of these participants in terms of age, or age of deafness diagnosis, coding or decoding (see data on these participants, number 2, 5 and 17, on Table

1). More detailed information on these participants provided by the experimenter establishes that two of these participants (2 and 17) are actually rather poor decoders, using decoding only for restricted tasks (participant 2 decodes seldom, and participant 17 decodes only for acquiring novel words). For the third participant (number 5) the performance is more surprising, since this participant does decode regularly, and it is likely that the possibly stressing situation provided by the experimental framework is responsible for the low performance. If we remove these 3 participants from the analysis of the relationship between SNR and CS gain, the Pearson coefficients actually increase (for AVC vs. A, $r^2 = 0.336$, $t(15) = 2.740$, $p = 0.015$); for AVC vs. AV, $r^2 = 0.359$, $t(15) = 2.901$, $p = 0.011$).

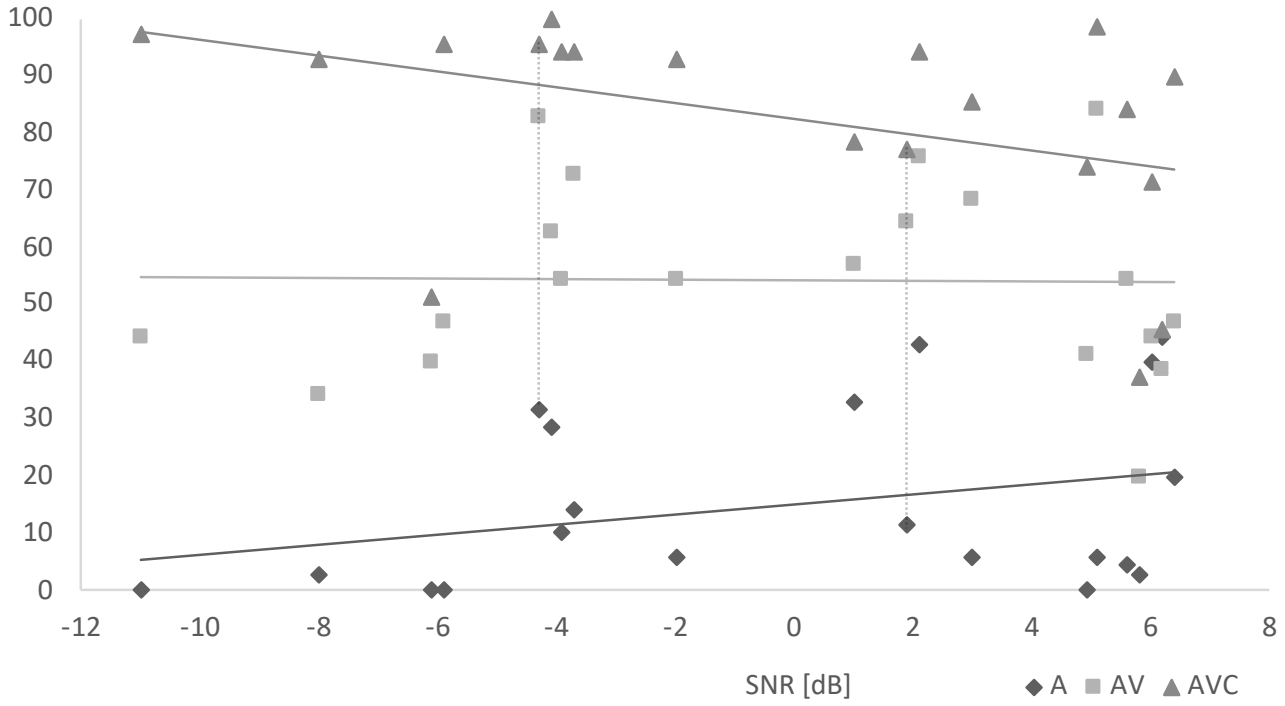


Figure 5. Variations of recognition scores in percent, in the three conditions (A, AV, AVC), as a function of SNR in the D/HH group. The plain lines display the linear regression fits to the data for each of the three conditions. To enable to assess performance of individual participants, light vertical dotted lines relate A, AV and AVC values for each participant (with slight modifications applied to SNR values in the figure to separate participants with identical SNR values in the experimental paradigm). The 3 circles display the 3 participants with poor AVC performance.

Discussion

The present study assessed speech perception in noise with visual information including both lip-reading and CS. It provides two main results. Firstly, in the absence of manual cues, there is a large

gap in reception of speech in noise between TH and D/HH participants, even with the addition of lip-reading. Secondly, CS significantly enhanced speech perception in noise for D/HH CS users. We will analyze these two points in more detail before discussing some clinical implications.

Differences in perception between typically hearing participants and participants with hearing loss fitted with CI or HA

The mean SNR providing around 60% reception in the AV condition, and around 15% in the A condition, is 11 dB higher in the D/HH group than in the TH group (Fig. 3). Since there is actually a large variability in all aspects of this pattern, let us describe it in more detail. The SNR values provide a 40-to-80% reception score in the AV condition for most (14 over 15) TH speakers and are distributed between -14 and -9 dB, while for the same AV score range (for 16 over 20 in the group) D/HH participants display an SNR between -11 and +6 dB. The SNR is actually above 0 dB for 10 D/HH participants (half the group). Finally, for 7 among the 20 D/HH participants (a third of the group) AV scores are actually lower than 60% for these SNR values above 0 dB, which makes intelligibility of speech at these SNRs quite poor.

Coming back to mean values, the average score in the A condition for the D/HH group is at 15% for an average SNR at 0dB. This is in line with previous studies. Indeed, Fu et al. (1998) conducted a series of experiments on the auditory recognition of vowels and consonants embedded in noise by typically hearing persons presented with stimuli simulating cochlear implants. The results show that at an SNR equal to 0 dB, recognition of vowels and consonants is perfect for non-degraded speech, while it decreases to less than 60% for speech spectrally compressed in 8 bands, which is compatible with very low recognition scores for words. Similar values are obtained by Friesen et al. (2001). A number of

444 other studies confirm that with an SNR of 0 dB audio perception is generally quite low in CI persons
445 (see e.g. Caldwell & Nittrouer, 2003; Tabanez do Nascimento & Bevilacqua, 2005).

446 The comparison of the A and AV conditions showed that the gain associated with lip-reading is similar
447 in the TH and D/HH groups around 40 percentage point (Fig. 4). This confirms that audiovisual fusion
448 does function efficiently in D/HH persons fitted with CI or HA as it has been found in a number of
449 studies (e.g., Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997; Lachs et al., 2001; Kaiser, Kirk,
450 Lachs, & Pisoni, 2003; Bergeson et al., 2005; Rouger et al., 2007). It could have been expected that
451 the AV gain would be higher for deaf persons considering that they put more weight on the visual input
452 (Desai et al. 2008) and that they are claimed to be better multisensory integrators (see Rouger et al.,
453 2007). However, a careful examination of the data in Rouger (2007) shows that, while there is indeed
454 a much higher visual gain for CI deaf participants than for typically hearing participants for spectrally
455 degraded speech ("vocoded speech"), there is much less difference for speech in noise (Rouger et al.
456 2007, see their Fig. 2) in line with the present study.

457 It is important to note that an SNR of 0 dB consists of noise with similar energy as the signal, which is
458 typically the case in a conversation with several partners. At such an SNR, audio-only word recognition
459 scores across languages usually reach more than 80% for TH persons (for English, see Cooke et al.
460 2013; for Spanish see Aubanel, García Lecumberri, & Cooke, 2014; for French see Aubanel et al.,
461 submitted). Hence, in comparison, the mean audio-only word recognition score of 15% for D/HH
462 participants in our study is extremely low. As mentioned previously, AV comprehension is also quite
463 low and does not ensure efficient comprehension for D/HH persons in noise conditions which are quite
464 frequent in everyday life. Understanding less than 60% of words means that global comprehension is
465 rather degraded, even if D/HH participants benefit from lip-reading. Considering that the ambient SNR

in classrooms may be typically as low as -6 dB (Picard & Bradley 2001), this illustrates the urgent need for complementary means to reach efficient communication for this population.

Benefits of Cued Speech

As argued above, Cued Speech might be an extremely important tool for D/HH persons. In fact, word recognition jumps by 28 percentage points (55% in AV to 83% in AVC) to a level of comprehension that renders communication possible (Fig. 4). In more detail, analysis of Fig. 5 shows that in the AVC condition, 17 among the 20 participants display word recognition scores above 75% at the tested SNR. Hence, the visual information provided by the combination of lip-reading and Cued Speech enables D/HH to recover comprehension to a level similar to scores displayed by TH participants who have no visual information, as displayed by audio-only word recognition scores in previous studies (see previous sub-section).

Note that in the AVC condition, we cannot separate the contributions of the various sources of information. As mentioned in the Introduction, it is known that lip-reading and manual cues suffice to reach a good level of comprehension (e.g. Nicholls & Ling, 1982). Hence, since we did not introduce a V + CS condition without sound, it is unclear to which extent acoustic information was used at all in the AVC condition. Still, since the experiment mixed all conditions within a single block, it is likely that participants keep taking profit of all the information available throughout the task. Indeed, the experimental data clearly show that audition is involved in the A condition, lip-reading does play a role in the AV condition and manual cues do intervene for CS readers in the AVC condition. Moreover, Bayard et al. (2014) have shown that deaf participants do integrate sound, lips and hands into a single percept. Finally, even if fusion per se was not tested in the present study, we can say at this stage that D/HH participants appear to be able to switch efficiently from A, AV to AVC conditions. In fact, this task needs to be solved frequently by D/HH CS users in conversations between typically hearing people

and CS users where some interlocutors are not directly visible, some are visible but not cueing, and some are visible and cueing. Importantly, our study shows the importance of the information conveyed by CS in this kind of situation.

The analyses of correlations between SNR and reception scores shed some interesting light on the potential efficiency of CS in communication for D/HH participants. Indeed, we found a significant correlation between SNR and gain in performance associated to the CS input, either alone (AVC-AV) or in combination with lip-reading (AVC-A). This interestingly shows that a high level of performance in CS decoding (evaluated by the AVC score or the related CS gains) does not impede a good performance without CS, quite on the contrary. It even seems that the best D/HH participants in terms of audiovisual speech reception in noise could be those who benefit most from the CS input. This is actually in line with previous studies by Leybaert & LaSasso (2010) or Aparicio, Peigneux, Charlier, Neyrat, & Leybaert, (2012) showing that Cued Speech provides a gain in audiovisual training enabling to improve speech perception in noise in D/HH persons. In any case, the important point is that CS is useful for improving speech perception in noise for D/HH persons mastering this process, and it does not seem to impede efficient A and AV perception without CS for these persons.

We did not find any benefit of CS for word recognition in typically hearing participants (53% for AV and 52% for AVC). We mentioned in Introduction that the CS system might enable the perceiver to focus attention in time. However, it shows that the temporal information leads to no reception benefit for those participants who do not know the phonetic interpretation of the hand positions and shapes. The reason is probably that there is already a good amount of timing information provided by the lip movements (see e.g. Grant & Seitz 1998, 2000; Kim & Davis 2014). Because of this redundancy and in the light of recent results showing a robustness of the temporal information benefit across a range of time delays (Aubanel, Masters, Kim, & Davis, 2017), there is probably no more room for

512 improvement with CS for non CS-users. Follow-up experiments might include acoustic stimuli with
513 noise that are accompanied by CS gestures without visible lips to control for this possibility that was
514 out of the scope of the present study. We are currently investigating this question by means of
515 electrophysiology.

516 Of course, this does not mean that only D/HH participants using CS should benefit from manual cues
517 in this task. It is expected that typically hearing persons using CS should display typically the same
518 gain between the AV and AVC conditions, unless there exists a specific advantage in the fusion of lip
519 and hand cues in D/HH persons, which to our knowledge has never been tested. The fact that the
520 analysis of score evolution along the experiment did not display any learning effect shows that the
521 experiment was too short, and quite probably too complex, to enable TH participants to detect some
522 specificities of manual cues that could have enabled them to improve their performance. This is in fact
523 unsurprising, considering the long time required for learning the CS system before efficient decoding
524 (see e.g. Clarke & Ling, 1976).

525 **Clinical implications**

526 Due to the technology progress, there is a trend that children with cochlear implants do not consistently
527 look at a speaker's mouth and hands (Marthouret, 2011). The consequence is that some parents may
528 lose their motivation to use Cued Speech, feel discouraged, or simply abandon coding with the hands
529 (Leybaert & Lassaso, 2010). In the light of our results, it appears relevant and important for D/HH
530 persons to maintain Cued Speech decoding abilities. Situations in quiet are rare in real life. Whether
531 this is in the personal or public sphere, background noises are pervasive. Accordingly, it would be
532 important for audiologists, speech therapists, educators, and related service providers to reflect
533 regularly on cueing necessity in certain contexts (e.g. periods when the child is tired, speech perception
534 in noisy situations etc.). Concerning re-education, the major challenge of speech therapists should be

535 to find the right equilibrium between the various sources of information, audio, labial and manual.
536 Focussing on auditory recovery and speech-reading is important to allow children to take full advantage
537 of their cochlear implant or hearing aid, but including CS in the re-education process might be of
538 importance for achieving efficient communication in specific situations, particularly involving noise
539 and adverse conditions.
540
541

542 ## Conclusion

543 The present study confirmed that speech perception in noise remains a challenge for D/HH persons
544 fitted with CI of HA. Importantly, for most D/HH participants, only the combination of audition, lip-
545 reading and manual cues enabled them to reach an adequate level of perception in noise (typically
546 above 80% correct words). Speech perception is a multimodal process in which different kinds of
547 information are likely to be merged: phylogenetically inherited phonetic information (provided by lip-
548 reading and audition) or recently invented additional relevant information (such as CS cues). Thus for
549 D/HH CS users fitted with cochlear implants, CI and CS could be a successful combination, in
550 particular in noisy environments allowing these persons to further improve their speech
551 comprehension.
552

553 **References**

554

555 Alegria, J., & Lechat, J. (2005). Phonological processing in deaf children : when lipreading and cues
556 are incongruent. *Journal of Deaf Studies and Deaf Education, 10*(2), 122-133.
557 doi:10.1093/deafed/eni013

558 Aparicio, M., Peigneux, P., Charlier, B., Neyrat, C., & Leybaert, J. (2012). Early experience of Cued
559 Speech enhances speechreading performance in deaf. *Scandinavian Journal of Psychology*, *41*,
560 41-46. doi: 10.1111/*j*.1467-9450.2011.00919.x

561 Aparicio M., Peigneux P., Charlier B., Balériaux D., Kavec M., & Leybaert J. (2017). The neural basis
562 of speech perception through lipreading and manual cues: evidence from deaf native users of
563 cued speech. *Frontiers in Psychology, 8*, 426. doi: 10.3389/fpsyg.2017.00426

564 Attina, V., Beautemps, D., Cathiard, MA., & Odisio, M. (2004). A pilot study of temporal organization
565 in Cued Speech production of French syllables: rules for a Cued Speech synthesizer. *Speech*
566 *Communication, 44*(1), 197-214. doi:10.1016/j.specom.2004.10.013

567 Attina, V. (2005). *La Langue française Parlée Complétée: production et perception.* Thèse de
568 Sciences Cognitives, Institut National Polytechnique de Grenoble-INPG.

569 Attina, V., Cathiard, MA., & Beautemps, D. (2006). Temporal measures of hand and speech
570 coordination during french cued speech production. In S. Gibet, N. Courty & J.F. Kamp (Eds.),
571 *Gesture in Human-Computer Interaction and Simulation* (Vol. 3881, pp. 13-24). Berlin,
572 Germany: Springer.

573 Aubanel, V., García Lecumberri, M. L., & Cooke, M. (2014). The Sharvard corpus: A phonemically-
574 balanced Spanish sentence resource for audiology. *International Journal of Audiology, 53*, 633–
575 638. doi: 10.3109/14992027.2014.907507

576 Aubanel, V., Masters, C., Kim, J., & Davis, C. (2017). Contribution of visual rhythmic information
577 to speech perception in noise. *Proceedings of the 14th International Conference on Auditory-*
578 *Visual Speech Processing (AVSP2017)*, Stockholm, Sweden.
579 Aubanel, V., Bayard, C., Strauß, A., & Schwartz, J.L. (submitted). The Fharvard corpus: A
580 phonemically-balanced French sentence resource for audiology.
581 Baer, T., Moore, B.C., & Gatehouse, S. (1993) Spectral contrast enhancement of speech in noise for
582 listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response
583 times. *Journal of Rehabilitation Research & Development, 30,* 49–72.
584 Bayard, C., Colin, C., & Leybaert, J. (2014) How is the McGurk effect modulated by Cued Speech in
585 deaf and hearing adults ? *Frontiers in Psychology. 5:416*. doi: 10.3389/ fpsyg.2014.00416
586 Bayard, C., Leybaert, J., & Colin, C. (2015) Integration of auditory, labial and manual signals in cued
587 speech perception by deaf adults : an adaptation of the McGurk paradigm. *Proceedings of The*
588 *1st Joint Conference on Facial Analysis, Animation and Auditory-Visual Speech Processing*
589 *FAAVSP-2015,* 163-168.
590 Bergeson, T.R., Pisoni, D.B., & Davis, R.A. (2005). Development of audiovisual comprehension
591 skills in prelingually deaf children with cochlear implants. *Ear and Hearing, 26(2),* 149-64.
592 Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (1998). What makes a good speechreader? First
593 you have to find one. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by Eye II. The*
594 *psychology of speechreading and auditory-visual speech* (pp. 211-228). East Sussex, U.K:
595 Psychology Press.
596 Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing.
597 *Perception & Psychophysics, 62 (2)*, 233-252.

598 Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in
599 fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the*
600 *Acoustical Society of America*, *125*, 3358–3372. doi: 10.1121/1.3110132
601 Bernstein, J.G., & Brungart, D.S. (2011) Effects of spectral smearing and temporal fine-structure
602 distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. *The*
603 *Journal of the Acoustical Society of America,* 130, 473–488. doi: 10.1121/1.3589440
604 Boothroyd, A., Mulhearn, B., Gong, J., & Ostroff, J. (1996) Effects of spectral smearing on phoneme
605 and word recognition. *The Journal of the Acoustical Society of America, 100,* 1807–1818. doi:
606 10.1121/1.416000
607 Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996. Intelligibility of normal speech I: Global and
608 fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255–272. doi:
609 10.1016/S0167-6393(96)00063-5
610 Bratakos, M. S., Duchnowski, P., & Braida, L. D. (1998). Toward the automatic generation of Cued
611 Speech. *Cued Speech Journal, 6*, 1-37.
612 Caldwell A., & Nittrouer, S. (2013). Speech perception in noise by children with cochlear implants.
613 *Journal of Speech, Language and Hearing Research, 56(1)*, 13-30. doi: 10.1044/1092-4388
614 Clark, B., & Ling, D. (1976) The effect of using cued speech : A follow-up study. *The Volta review,*
615 *78(1),* 23-34.
616 Cooke M., Mayo C., Valentini-Botinhao C., Stylianou Y., Sauert B. et al. (2013). Evaluating the
617 intelligibility benefit of speech modifications in known noise conditions. *Speech*
618 *Communication, 55*, 572–585. Doi : 10.1016/j.specom.2013.01.001
619 Cornett, R. O. (1967). Cued speech. *American annals of the deaf, 112*(1), 3-13.
620 Cornett, R. O. (1994). Adapting Cued Speech to additional languages. *Cued Speech Journal, 5*, 19-
621 29.

622 Duchnowski, P., Lum, D.S., Krause, J.C., Sexton, M.G., Bratakos, M.S., & Braida, L.D. (2000).
623 Development of speechreading supplements based on automatic speech recognition. IEEE
624 Transactions *on* Biomedical Engineering*, 47(4),* 487-96.
625 Erber, N. P. (1971). Auditory and audiovisual reception of words in low-frequency noise by
626 children with normal hearing and by children with impaired hearing. *Journal of Speech and*
627 *Hearing Disorders, 14*, 496-512 .
628 Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders,*
629 *40*(4), 481-492.
630 Friesen, L.M., Shannon, R.V., Baskent, D., & Wang, X. (2001) Speech recognition in noise as a
631 function of the number of spectral channels: Comparison of acoustic hearing and cochlear
632 implants. *The Journal of the Acoustical Society of America, 110,* 1150–1163. doi:
633 10.1121/1.1381538
634 Fu, Q.J., Shannon, R.V., & Wang, X. (1998) Effects of noise and spectral resolution on vowel and
635 consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of*
636 *America, 104,* 3586–3596. doi: 10.1121/1.423941
637 Gibert, G., Bailly, G., Beautemps, D., Elisei, F., & Brun, R. (2005). Analysis and synthesis of the
638 three-dimensional movements of the head, face, and hand of a speaker using cued speech. *The*
639 *Journal of the Acoustical Society of America*, *118(2)*, 1144-53. doi: 10.1121/1.1944587
640
641 Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-
642 impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration.
643 *The Journal of the Acoustical Society of America,103*, 2677–2690. doi: 10.1121/1.422788
644 Gregory, J. F. (1987). An investigation of speechreading with and without Cued Speech. *American*
645 *annals of the deaf, 132*(5), 393-398.

646 Holt, R. F., Kirk, K. I., & Hay-McCutcheon, M. (2011). Assessing multi-modal spoken word-in-
647 sentence recognition in children with normal hearing and children with cochlear implants.
648 *Journal of Speech, Language, and Hearing Research*, 54*, 632–657.*
649 Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual
650 word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing*
651 *Research, 46*, 390–404.
652 Kim, J., & Davis, C. (2014). How visual timing and form information affect speech and non-speech
653 processing. *Brain and Language*, *137*, 86–90. doi: 10.1016/j.bandl.2014.07.012.
654 Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by
655 prelingually deaf children with cochlear implants: A first report. *Ear and Hearing, 22*, 236–251.
656 Leybaert, J., & LaSasso, C. J. (2010). Cued speech for enhancing speech perception and first language
657 development of children with cochlear implants. *Trends in Amplification 14(2)*, 96-112. doi:
658 10.1177/1084713810375567
659 Liu, Shu-Yu, Yu, Grace, Lee, Li-Ang, Liu, Tien-Chen, Tsou, Yung-Ting, Lai, Te-Jen, & Wu, Che-
660 Ming. (2014). Audiovisual Speech Perception at Various Presentation Levels in Mandarin-
661 Speaking Adults with Cochlear Implants. *PloS ONE, 9*. e107252. doi:
662 10.1371/journal.pone.0107252.
663 Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. (2006). Speech perception problems of
664 the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National*
665 *Academy of Sciences U S A, 103,* 18866-18869. doi: 10.1073/pnas.0607364103
666 Marthouret, M. (2011). Faut-il proposer la Langue Parlée Complétée à des sourds porteurs d'un
667 implant cochléaire? (Should we propose Cued Speech to deaf children with a cochlear implant?)
668 In J. Leybaert (Ed.): *La langue Parlée Complétée: Fondements et Perspectives (Cued Speech:*
669 *Foundations and Future)*. Marseille, France: Solal.

Nicholls, G. H., & Ling, D. (1982). Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research, 25*(2), 262-269.

Percy, V., Raymond, B., Smith, A., Joseph, C., Kronk, L., Henry, B., & Kei, J. (2013). Measuring speech perception abilities in adults with cochlear implants: Comprehension versus speech recognition. *Australian and New Zealand Journal of Audiology*, *33(1),* 35-47.

Périer, O., Charlier, B., Hage, C., & Alegria, J. (1990). Evaluation of the effect of prolonged Cued Speech practice upon the reception of spoken language. *Cued Speech J. IV,* 45–59.

Picard, M., & Bradley, J.S. (2001). Revisiting speech interference in classrooms. *Audiology, 40*, 221-244.

Revoile, S.G., Pickett, J.M., & Kozma-Spyteck, L. (1991) Spectral cues to perception of/d, n, l/ by normal- and impaired-hearing listeners. *The Journal of the Acoustical Society of America, 90,* 787–798. doi/10.1121/1.401948

Rothauser, E. H., Chapman, W. D., Guttman , N., Hecker M. H. L., Nordby K. S. et al. (1969). IEEE Recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 225–246. doi: 10.1109/IEEESTD.1969.7405210

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (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. doi: 10.1073/pnas.0609419104

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory–visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 18748–18750. doi: 10.1073/pnas.0508862102

692 Schwartz, J.-L., & Savariaux, C. (2014) No, there is no 150 ms lead of visual speech on auditory
693 speech, but a range of audiovisual asynchronies varying from small audio lead to large audio
694 lag. *PLoS Computational Biology, 10(7),* e1003743. doi: 10.1371/journal.pcbi.1003743
695 Shannon, R.V., Fu, Q.-J., & Galvin, J. III. (2004). the number of spectral channel required for speech
696 recognition depends on the difficulty of the listening situation. *Acta Otolaryngol, Suppl 552,* 50-
697 54.
698 Shannon, R.V., Fu, Q.-J., & Galvin, J. III., & Friesen, L. (2004). Speech perception with cochlear
699 implants. In F.G. Zeng, A.N. Popper, & R.R. Fay (Eds.) *Auditory Prostheses and Electric*
700 *Hearing. Springer Handbook of Auditory Research, vol 20* (pp 334-376). New-York, NY:
701 Springer.
702 Srinivasana, A. G., Padilla, M., Shannon, R. V., & Landsberger, D. M. (2013). Improving speech
703 perception in noise with current focusing in cochlear implant users. *Hearing Research, 299*, 29–
704 36. doi: 10.1016/j.heares.2013.02.004
705 Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech:
706 revisiting the N400. *Journal of Cognitive Neuroscience*, 25(8), 1383-95. doi:
707 10.1162/jocn_a_00389
708 Tabanez do Nascimento, L., & Bevilacqua, M. C. (2005). Evaluation of speech perception in noise in
709 cochlear implanted adults. *Revista Brasilera Otorrinolaringolia,* 71(4), 432-438. doi : S0034-
710 72992005000400006
711 Taitelbaum-Swead, R., & Fostick, L. (2017). Audio-visual speech perception in noise: Implanted
712 children and young adults versus normal hearing peers. *International Journal of Pediatric*
713 *Otorhinolaryngology, 92,* 146-150. doi: 10.1016/j.ijporl.2016.11.022

714 Todorov, M.T., & Galvin,K.L. (2018). Benefits of upgrading to the nucleus® 6 sound processor for
715 a wider clinical population. *Cochlear Implants International*, *19(4),* 210-215. doi:
716 10.1080/14670100.2018.1452584
717 Troille, E., Cathiard, MA., & Abry, C. (2007). A perceptual desynchronization study of manual and
718 facial information in French Cued Speech. *ICPhS, Saarbrücken, Germany*, 291-296.
719 Tutz, G., & Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational*
720 *Statistics & Data Analysis, 22*, 537–557. doi: 10.1016/0167-9473(96)00004-7
721 Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., & Gantz, B. J. (1997). Performance
722 over time of adult patients using the Ineraid or Nucleus cochlear implant. *Journal of the*
723 *Acoustical Society of America, 102*, 508–522. doi/10.1121/1.419724
724 Uchanski, R. M., Delhorne, L. A., Dix, A. K., Braida, L. D., Reed, C. M., & Durlach, N. I. (1994).
725 Automatic speech recognition to aid the hearing impaired: prospects for the automatic
726 generation of cued speech. Journal of Rehabilitation Research and Development, 31(1), 20-41.
727 Wolfe, J., Morais, M., Schafer, E., Agrawal, S., & Koch, D. (2015). Evaluation of speech recognition
728 of cochlear implant recipients using adaptive, digital remote microphone technology and a
729 speech enhancement sound processing algorithm. *Journal of the American Academy of*
730 *Audiology, 26*, 502–508. doi: 10.3766/jaaa.14099
731
732 Zeng, F. G., & Galvin, J. (1999). Amplitude mapping and phoneme recognition in cochlear implant
733 listeners. *Ear & Hearing, 20*, 60-74.
734

735
736 **Table 1. Characteristics of deaf participants (CI: cochlear implant; HA: hearing aid)**

737 To ensure confidentiality, age is provided only in ranges of 3-years. Deafness degree, cause of
738 deafness, age of diagnosis, age of cueing (i.e. age of expressive cueing) and decoding (i.e. age of
739 receptive cueing), age of left and right ear equipment, were communicated by the participants. ??
740 indicates that information since when CI or HA where used is unknown. SNR level was determined
741 individually in a pre-test (see text). Participants are ordered by decreasing values of their SNR.

Parti-cipant	Age range (years)	Gender	Age of diagnosis (months)	Deafness degree	Cause of deafness	Age of cueing (years)	Age of decoding (years)	Age of left ear equipment	Age of right ear equipment	SNR level (dB)
1	12-14	F	At birth	Profound	Connexine 26	7,5	3,5	CI (6 yr)	CI (2 yr)	6
2	14-16	F	13	Profound	Unknown	Unknown	1,5	HA (birth)	CI (2,5 yr)	6
3	12-14	F	20	Profound	Unknown	4	4	HA (2,5 yr)	HA (2,5 yr)	6
4	18-20	M	8	Profound	Unknown	8	8	None	CI (5 yr)	6
5	12-14	M	9	Profound	Waardenburg syndrome	2	1	CI (1 yr)	CI (1yr)	6
6	16-18	M	12	Profound	Unknown	7	6	CI (3 yr)	None	5
7	20-22	F	9	Profound	Unknown	7	5,5	CI (7 yr)	HA (1 yr)	5
8	14-16	F	20	Profound	Unknown	4	5	HA (??)	CI (11 yr)	3
9	16-18	M	18	Profound	Unknown	3	3	None	CI (6 yr)	2
10	20-22	F	6	Severe	Unknown	5,5	6	CI (4 yr)	None	2
11	12-14	M	11	Profound	Connexine 26	8	3	CI (12 yr)	CI (2 yr)	1
12	14-16	M	18	Profound	Genetic	Unknown	4	CI (4yr)	HA (1.5 yr)	-2
13	12-14	F	48	Profound	Unknown	7	6	HA (??)	HA (??)	-4
14	12-14	M	12	Profound	Connexine 26	1,5	3	CI (9 yr)	HA (??)	-4
15	12-14	M	18	Profound	Cytomegalo-virus	7	2	CI (??)	CI (??)	-4
16	20-22	M	30	Profound	Otitis	5	5	CI (12 yr)	CI (18 yr)	-4
17	14-16	F	15	Profound	Connexine 26	9		HA (??)	CI (15 yr)	-6
18	20-22	M	12	Profound	Unknown	5,5	2	HA (1 yr)	HA (1 yr)	-6
19	14-16	M	9	Profound	Pendred syndrome	7	2	CI (5 yr)	CI (15 yr)	-8
20	16-18	F	9	Profound	Connexine 26	4	1	CI (16 yr)	HA (1 yr)	-11

742

743 **Table 2. Characteristics of typically hearing participants**

744 SNR was determined individually in a pre-test (see text). Participants are ordered by decreasing

745 values of their SNR.

746

Participant	Age range (years)	Gender	SNR level (dB)
1	30-32	F	-9
2	26-28	M	-9
3	24-26	F	-9
4	30-32	F	-10
5	30-32	M	-10
6	28-30	F	-10
7	22-24	M	-11
8	34-36	F	-11
9	28-30	F	-11
10	36-38	F	-12
11	28-30	M	-12
12	30-32	M	-13
13	28-30	F	-13
14	26-28	F	-14
15	32-34	M	-14

747

748

749 **Table 3. Experimental design. Pseudo-randomization procedure.**
750

Conditions		Number of sentences	Order 1	Order 2
AV		14	Sent. 1 to 14	Sent. 15 to 28
AVC		14	Sent. 15 to 28	Sent. 1 to 14
A	Sound from AV	7	Sent. 29 to 35	Sent. 36 to 42
	Sound from AVC	7	Sent. 36 to 42	Sent. 29 to 35

751
752