Degradation of Labial Information Modifies Audiovisual Speech Perception in Cochlear-Implanted Children

Aurélie Huyse,[1] Frédéric Berthommier,[2] and Jacqueline Leybaert[1]

Objective: The aim of the present study was to examine audiovisual speech integration in cochlear-implanted children and in normally hearing children exposed to degraded auditory stimuli. Previous studies have shown that speech perception in cochlear-implanted users is biased toward the visual modality when audition and vision provide conflicting information. Our main question was whether an experimentally designed degradation of the visual speech cue would increase the importance of audition in the response pattern. The impact of auditory proficiency was also investigated.

Design: A group of 31 children with cochlear implants and a group of 31 normally hearing children matched for chronological age were recruited. All children with cochlear implants had profound congenital deafness and had used their implants for at least 2 years. Participants had to perform an /aCa/ consonant-identification task in which stimuli were presented randomly in three conditions: auditory only, visual only, and audiovisual (congruent and incongruent McGurk stimuli). In half of the experiment, the visual speech cue was normal; in the other half (visual reduction) a degraded visual signal was presented, aimed at preventing lipreading of good quality. The normally hearing children received a spectrally reduced speech signal (simulating the input delivered by the cochlear implant).

Results: First, performance in visual-only and in congruent audiovisual modalities were decreased, showing that the visual reduction technique used here was efficient at degrading lipreading. Second, in the incongruent audiovisual trials, visual reduction led to a major increase in the number of auditory based responses in both groups. Differences between proficient and nonproficient children were found in both groups, with nonproficient children's responses being more visual and less auditory than those of proficient children. Further analysis revealed that differences between visually clear and visually reduced conditions and between groups were not only because of differences in unisensory perception but also because of differences in the process of audiovisual integration per se.

Conclusion: Visual reduction led to an increase in the weight of audition, even in cochlear-implanted children, whose perception is generally dominated by vision. This result suggests that the natural bias in favor of vision is not immutable. Audiovisual speech integration partly depends on the experimental situation, which modulates the informational content of the sensory channels and the weight that is awarded to each of them. Consequently, participants, whether deaf with cochlear implants or having normal hearing, not only base their perception on the most reliable modality but also award it an additional weight.

(Ear & Hearing 2013;34:110–121)

INTRODUCTION

In face-to-face communication, speech perception is a multimodal process involving both auditory and visual modalities (Sumby & Pollack 1954; Grant & Seitz 2000). Studies have shown that speech detection and comprehension are better in audiovisual (AV) conditions, where audition is accompanied by lipreading, than in auditory-only (AO) conditions, where only the auditory stimulus is present (Erber 1969; MacLeod & Summerfield 1987; Grant & Seitz 2000). This is because of the fact that during speech perception, auditory and visual cues are merged into a unified percept, a mechanism known as audiovisual integration. AV integration is illustrated by the McGurk effect (McGurk & MacDonald 1976) that occurs when audition and vision provide incongruent tokens. For example, when presented with visual velar /ka/ and auditory bilabial /pa/, normally hearing individuals (NH) tend to report the illusory fusion alveo-dental /ta/. The McGurk fusions show that visual articulatory cues about place of articulation (lipreading) are integrated into the auditory percept, which is, thereby modified.

The visual modality exerts even a greater influence on speech perception in some groups, such as cochlear-implanted (CI) deaf people. At first, this neuroprothesis was designed as a single-electrode device aiming to enhance lipreading and provide sound awareness (Zeng 2004). Today, it is a sophisticated multi-electrode hearing device helping deaf people achieve high levels of speech recognition. For example, postlinguistically deafened adults achieve almost 90% correct identification of words in quiet surroundings (Fetterman & Domico 2002). Nevertheless, the stimulation provided to the auditory nerve by the cochlear implant (CI) is different from the normal stimulation, resulting in modified information, particularly regarding the place of articulation, voicing, and nasality cues (Dowell et al. 1982; Skinner et al. 1999; Kiefer et al. 2001). For example, the information relative to place of articulation coding delivered through the CI is not accurate enough to distinguish words that differ only by the place of articulation, as "buck/duck." Thereby, the need for lipreading is increased (Giraud et al. 2001), although many important components of speech articulation are not directly visible under normal circumstances (Mohammed et al. 2005). Consequently, AV speech perception in CI users is a particularly interesting topic to investigate.

Several studies have shown that postlinguistically deafened adults with CI are able to integrate congruent auditory and visual cues appropriately, showing an advantage in the AV modality compared with the AO modality (visual gain) (Tyler et al. 1997; Kaiser et al. 2003). Moody-Antonio et al. (2005) showed that even some prelinguistically deafened subjects who were implanted as adults had better performances in AV modality than for unimodal AO and visual-only (VO) modalities. It has also been suggested that postlinguistically deafened adults with CI are better AV integrators than NH individuals. Using an AV speech-recognition task, Rouger et al. (2007) found that CI users had higher AV performances and higher visual gains than NH participants presented with noise-band vocoder speech did, although both groups had similar auditory levels. However, differences in unimodal conditions might, in fact, be responsible for that result. Schwartz (2010) pointed out that the

---

[1] Department of Psychological Science, Laboratoire Cognition, Langage, Développement, Université Libre de Bruxelles, Belgium; and [2] GIPSA-Lab, Grenoble University, France.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text versions of this article on the journal's Web site (www.ear-hearing.com).

0196/0202/12/3401-0110/0 • Ear & Hearing • Copyright © 2012 by Lippincott Williams & Wilkins • Printed in the U.S.A.

transmission of the voicing mode was poorer in speech degraded with noise-band vocoder, suggesting that there could be a poorer complementary in the audio and visual inputs in NHV, logically resulting in lower audiovisual scores. Differences in audiovisual performance would hence result from the structure of the unimodal inputs being less complementary for normal hearing subjects presented with noise-band vocoded speech rather than from integration per se (p. 1592).

Nevertheless, the abovementioned studies emphasize that adults with CI have specific visual and audiovisual skills, which allow them to compensate for the difficulties related to the CI.

Two recent studies investigating the responses to incongruent AV stimuli found that, when faced with McGurk stimuli, CI deaf adults in comparison to NH controls tend to favor the visual modality (Desai et al. 2008; Rouger et al. 2008). It is probable that the importance of vision to CI patients is increased because of stronger lipreading skills developed during preimplantation deafness and because of the need to rely on vision to overcome the uncertainty of the auditory signal. Lipreading skills of CI users might also be enhanced after the implantation, as a result of mutual reinforcement of hearing speech and seeing speech (Giraud et al. 2001). Recently, Tremblay et al. (2010) explicitly investigated the link between CI proficiency and responses to incongruent McGurk stimuli in postlinguistically deafened adults. They divided the CI participants into two groups based on performance at a speech-recognition task in AO modality: The proficient (performances > 70% of correct identification) and the nonproficient (performances < 70% of correct identification). First, with an incongruent AV stimulus composed of an auditory /ba/ and a visual /ga/, they found that the amount of McGurk fusion /da/ in both groups was similar to that of a group of NH controls, revealing normal AV fusion abilities. Second, when another response than /da/ was made, the proficient group and the NH group tended to give a response that matched the auditory cue (/ba/) whereas the nonproficient favored the visual speech cue (/ga/). These data suggest that AV speech perception of CI users is directly related to CI proficiency.

Findings for AV speech perception in CI children are similar to those of CI adults. First, several studies showed that CI children with prelinguistic deafness perform better in congruent AV modality than in auditory modality, demonstrating a significant visual gain (Lachs et al. 2001; Geers et al. 2003; Bergeson et al. 2005). Second, a study in French using McGurk stimuli found that CI children produced mainly visually based responses to incongruent AV stimuli, whereas age-matched NH children tended to give auditory based responses (Leybaert & Colin 2007). In a study in English, Schorr et al. (2005) showed that the majority of NH children (20 of 35) made consistent fusion (they reported the fusion response on ≥ 7 of the 10 trials), whereas the majority of CI children (24 of 30) made visually based responses. A minority (6 of 30) of CI children made consistent fusion. Those children represented 38% of the children implanted before 2.5 years of age. Given that no CI child implanted after 2.5 years made consistent fusion, the authors suggested that AV integration of speech depends on AV language experience in early life (before the age of 2.5 years).

Taken together, these studies suggest that the AV integration of incongruent cues in CI users (postlinguistically deafened adults and prelinguistically deafened children) is atypical, leading to a natural AV imbalance in favor of vision. However, although AV imbalance depends on CI proficiency in postlinguistically deafened adults, the relationship between CI proficiency and McGurk fusion effect in CI prelinguistically deafened children remains to be investigated.

There is compelling evidence that, even in NH listeners, the McGurk fusion relies on interindividual factors such as language background (e.g., Sekiyama & Tokhura 1993) and age (e.g., Sekiyama et al. 2003; Sommers et al. 2005). Moreover, differences in McGurk fusion could also rely on intraindividual factors and be a result of the property of the task or the stimuli itself (Schwartz 2010). This suggests that the effect is not as automatic and irrepressible as has been thought for many years (Manuel et al. 1983; Liberman & Mattingly 1985; Rosenblum & Saldaña 1996). The results of a previous study including 40 NH adults support this view (Huyse et al. in revision). They performed a syllable-identification task in which targets were presented in four modalities: AO, VO, congruent AV, and incongruent AV (McGurk). Syllables were embedded in noise that was either stationary or modulated in amplitude (Füllgrabe et al. 2006). Participants were divided into two groups. In the control group, the visual signal was not manipulated; the video was clear. In the visual-reduction group, the quality of the visual signal was degraded by a technique of contrast modulation, to prevent lipreading of good quality. Results of the McGurk stimuli were the following: in the stationary noise, the control group gave mainly visually based responses; in the modulated noise they gave mainly McGurk fusions. Compared with the control group, the reduction group made significantly more auditory based responses, whether the noise was stationary or modulated. In other words, it seems that responses to incongruent AV stimuli rely strongly on the experimental situation and on the information contained in each modality.

The goal of the present research was to examine the importance of intraindividual factors in AV speech integration in CI children. More specifically, the present study aimed to determine whether a visual bias in CI children's response to McGurk stimuli can be decreased in favor of audition. To answer that question, 31 prelinguistically deafened children with CIs were presented with an AV speech-identification paradigm in which stimuli (six syllables containing voiceless consonants) were presented in four modalities: AO, VO, congruent AV, and incongruent AV (McGurk). During half of the experiment, the visual speech cue was clear (VCL) whereas it was degraded in the other half (visual reduction [VR]) to prevent participants from performing lipreading of good quality. If AV speech is indeed dependent on a weighting process, itself depending on the experimental situation, the weight of audition should be increased in the VR condition compared with the VCL condition. In other words, we expect CI children to give visually based responses in VCL (because of the visual bias) and more auditory based responses in VR. Moreover, performances of proficient CI users were compared with performances of nonproficient CI users to determine whether responses also depended on CI proficiency, as is the case for postlinguistically deafened adults (Tremblay et al. 2010).

The second goal of the present research was to compare the performance of CI children to that of NH children and to study the possibility of modifying the pattern of response of NH children placed temporarily in a hearing-degradation situation, with no training whatsoever. To that aim, a group of NH children participated in the study. In that group, the syllables transmitted in

AO and AV modalities were degraded by a spectral reduction of speech (SRS) paradigm that simulates the processing of the CI (Shannon et al. 1995). In SRS, the spectral information is degraded while the temporal information is preserved. It leads to a partial suppression of information on the place of articulation (Berthommier 2001). Multiple techniques exist to create a spectrally reduced speech signal, among which is the technique of Shannon et al. (1995). In their study, the spectrum was divided into a variable number of sub-bands from one to four. They found that the transmission of phonetic information was restored at four sub-bands, except for the place of articulation. In her doctoral dissertation, Grosgeorge (Reference Note 1) also used a variable number of sub-bands (1, 2, 4, and 16) and confirmed that information regarding the place of articulation depends on the spectral resolution of the speech signal. Indeed, with 4 sub-bands, the place of articulation was transmitted at a rate of 20%, whereas with 16 sub-bands the rate was 65%. SRS is therefore a good way to place NH individuals in situations similar to those of CI users as regards the perception of the place of articulation. For that reason, the SRS paradigm is a good alternative to a noise paradigm (as was used by Huyse et al. in revision) and was used here to compare both groups of participants. Moreover, performances of proficient NH children were compared with those of nonproficient NH children, as with CI users.

PARTICIPANTS AND METHODS

Participants

Thirty-one CI children (17 girls and 14 boys, mean age = 10 years, SD = 0.47) and 31 NH children (15 girls and 16 boys, mean age = 10 years, SD = 0.5) participated in this study. They were all native French speakers with normal or corrected-to-normal vision and did not have any language or cognitive disorders. All CI children had congenital profound deafness, had received their implant before 8 years of age, and had used it for at least 2 years. They all had a unilateral implant and seven children had contralateral hearing aids that were turned on during the testing session. Table 1 provides a summary of the main characteristics of the CI children.[a]

Stimuli Material

A male French speaker was videotaped while saying vowel–consonant–vowel (aCa) syllables. The cVc stimuli used in this study consisted of the consonants /p, t, k, s, f, ʃ/ coarticulated with two /a/ vowels. Three productions of each /aCa/ stimulus were digitally recorded and audio tracks were equalized in level. The speaker was filmed and viewed from the bottom of the nose to the chin. The production of each stimulus began and ended in a neutral, closed-mouth position for a total duration of about 300 msec. Videos were displayed centered on a 15-in monitor on a black background. The congruent AV stimuli included digital audio–video files of the speaker saying and articulating the aCa stimuli. For the AO condition, an image of the speaker, seeming neutral and with mouth closed was presented along with the auditory stimulus. For the VO condition, the audio was turned off. The six phonemes used in our study correspond to six different visemes (Walden et al. 1977) allowing a success rate above chance level.

TABLE 1. Characteristics of the children from the cochlear implant group

Participants	Age (yrs)	Age at Implantation (mos)	Duration Since Implantation (yrs)
1	8	13	7
2	10	13	9
3	11	18	10
4	7	24	5
5	9	24	7
6	10	26	8
7	14	27	12
8	8	28	6
9	9	28	7
10*	8	30	7
11	10	30	8
12	10	30	8
13	11	32	8
14	10	36	7
15	16	36	13
16*	11	36	8
17*	7	36	4
18	7	36	4
19	10	36	7
20	9	39	6
21	8	42	5
22	14	48	10
23	7	48	3
24	12	54	8
25*	7	60	2
26	9	60	4
27*	10	60	5
28*	17	72	11
29*	9	81	2
30	16	81	9
31	11	96	3

**Children who wear contralateral hearing aid.*

Finally, incongruent AV syllables (McGurk stimuli) were created by carefully combining audio files with noncorresponding video files and matching their onset. We used three repetitions of the two following stimuli: the audio /apa/ with the visual /aka/ (fusion /ata/) and the audio /afa/ with the visual /aʃa/ (fusion / asa/). Stimuli were delivered through Sennheiser HD 121 Pro headphones for the control group and through two loudspeakers for the CI group, one on each side of the computer screen.

The total number of items was 132 “visual-clear” stimuli and 132 “visual-reduction” stimuli (6 Syllables × 3 Repetitions × 3 Modalities + 12 McGurk stimuli, randomly mixed, repeated two times). Three blocks of 44 visual-clear items and 3 blocks of 44 visual-reduction items were created. For the visual-reduction stimuli, we used a technique of contrast modulation at a rate of 4 Hz. We alternated one visual-reduction block with one visual-clear block. In both groups half the participants began the experimental session with a visual-reduction block and the other half began with a visual-clear block. The task was to identify the syllable and to report it aloud.

Visual Reduction

To reduce the quality of the visual component, we varied the contrast of the video around the mean intensity of the image X, for each red, green, blue color of the image.

---

[a] The majority of the CI children were implanted before 5 years of age. However, removing those children from the sample did not vary significantly the scores in AO, VO, and AV modalities.

$$Y = \text{mean}(X) + \left(\frac{(X - \text{mean}(X))}{R}\right)$$

The contrast R varies at a period of 4 Hz according to the following function:

$$R = k \times 10^{\left(1+0.5*\cos\left(\phi + f(t)\right)\right)}$$

where the parameter k is set to 4 and where $f(t)$ represents the modulation frequency. $f(t)$ was set at 4 Hz. Because the total duration of the syllable (the mouth movements) exceeded 250 msec, a contrast modulation of 4 Hz always generated periods of masking and periods of unmasking within a single item. The starting phase of the modulation φ was randomized in each interval between 0 and 2 π. Because φ was random, VR did not affect each exemplar of a single phoneme equally.

Spectrally Reduced Speech Signal

The NH group received a spectrally reduced speech signal. To perform the SRS, a technique similar to that used in the study of Shannon et al. (1995) was applied, except that instead of using 4 sub-bands, we chose to use 16 sub-bands on the basis of two preliminary experiments. In the first preliminary experiment, 52 young NH adults were divided into four groups receiving different numbers of sub-bands: 2, 4, 8, and 16. They were tested in the same AV syllable-identification task as in this study. Their auditory performances, averaged on VCL and VR blocks, were compared with those of our CI group. We found that only the group that received 16 sub-bands had global auditory performances (82%, averaged over the 6 phonemes) that were not significantly different from those of the CI groups (76%, F[1, 60] = 1.86, p = 0.142) (Fig. 1). In the other pilot study, a group of 13 NH children (matched to the CI children for chronological age) performed the same task with eight sub-bands. We found that their auditory level (53%) was significantly lower than that of CI children, which supported our choice of 16 sub-bands [F(1, 42 = 32.86, $p < 0.005$].

Procedure

The experiment took place in a dimly lit, quiet room. Stimuli were presented on a monitor positioned at eye level, 70 cm from the participant's head. Participants were informed about the composition of the stimulus set (but not of the presence of McGurk stimuli) and had the response options in front of them during the training session but not during the experimental session. Children were asked to repeat the syllable perceived (heard or lipread). Before the testing, the experimenter asked each child to repeat each stimulus after him, to detect potential articulation disorders. The testing always began with a training session consisting of one visual-clear block and one visual-reduction block. Responses were noted by the experimenter. The total duration of the experiment was approximately 30 min.

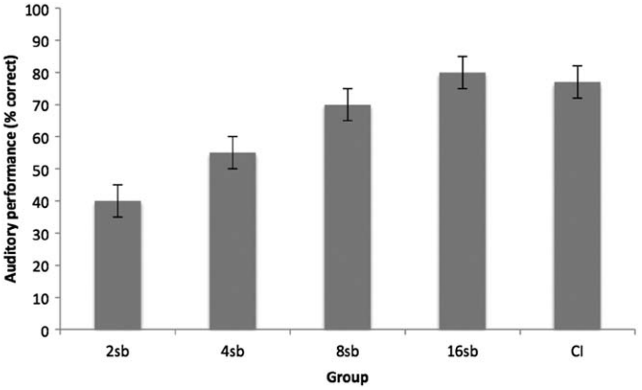

Fig. 1. Mean scores in auditory-only condition of the CI group and of NH young adults receiving a spectral reduced speech signal with 2, 4, 8, and 16 sub-bands. CI, cochlear implanted; NH, normally hearing.

RESULTS

Single and AV Congruent Modality Conditions

Table 2 shows speech-identification scores (percent correct and standard deviation) for each group. A mixed analysis of variance (ANOVA) was performed with modality (AO, VO, and AV), visual condition (VCL, VR), and group (CI, NH) as factors. The main effect of modality was significant (F[2, 120] = 246.62, $p < 0.0001$). Planned comparison with a Bonferroni correction showed that participants performed better in AV modality than in AO ($p < 0.0001$), in AV than in VO ($p < 0.0001$) and in AO than in VO ($p < 0.0001$). There was also a main effect of visual condition (F[1, 60] = 92.03, $p < 0.0001$); better performances were obtained in VCL compared with VR blocks. There was also a significant interaction between visual condition and modality (F[2, 120] = 129.10, $p < 0.0001$), showing that VR led to an increase in the AO modality ($p < 0.0001$) and to a decrease in the AV ($p = 0.001$) and VO ($p < 0.0001$) modalities. The main effect of group was not significant ($p = 0.35$) nor was any interaction involving the group factor.

The standardized AV gain (AVG) was calculated in VCL and in VR blocks using the following formula: (AV–AO)/(100–AO), where AO and AV are percentage recognition scores. AVG measures the relative increase in AV speech-perception performance because of the addition of visual information to the auditory signal (Sumby & Pollack 1954). An ANOVA was run with visual condition and group as factors. There was a main effect of visual condition (F[1, 60] = 20.68, $p < 0.0001$), but there was no main effect of group ($p = 0.18$) and no significant interaction between group and visual condition ($p = 0.37$). These results show that AVG was significantly decreased by VR and that this decrease was similar in both groups.

McGurk Effect

Results of the two McGurk stimuli used in our study were analyzed separately. The reason not to average them is that identification of the stimuli constituting the McGurk trials in AO and VO were significantly different. For example, in the AO modality,

TABLE 2. Global performance (% correct) of the cochlear implanted and normally hearing children

		Cochlear-Implanted Group	Normally Hearing Group
Auditory-only	VCL	76.16 (2.47)	79.66 (1.79)
	VR	80.29 (2.73)	85.04 (1.92)
Visual-only	VCL	63.26 (2.78)	64.96 (1.77)
	VR	36.65 (2.84)	42.03 (3.41)
Audiovisual	VCL	92.03 (1.66)	87.90 (1.70)
	VR	84.41 (2.60)	86.29 (1.93)
Audiovisual gain (AV–AO/100–AO)	VCL	64.85 (6.40)	36.68 (8.29)
	VR	7.06 (13.24)	0.85 (15.56)

VCL, visual clear; VR, visual reduction; AV, audiovisual; AO, audio only.

/afa/ was correctly identified at 59% whereas /apa/ was correctly identified at 71%. This difference between AO/afa/ and AO/apa/ was even more pronounced in the NH group (50% and 93%, respectively). Similarly, in the VO modality, percentage correct of identification of /aʃa/ was very different from the percentage correct of /aka/ (84% versus 66% in the CI group and 90% versus 59% in the NH group). Because identification of /apa/, /afa/, /aʃa/ and /aka/ in unimodal modalities differed significantly, one might expect performance in AV incongruent to be different as well. Thereby, responses to A/apa/V/aka/ and /A/afa/V/aʃa/ were analyzed separately instead of being averaged.

Scores consisted of proportions and were not normally distributed. Therefore, data were transformed using an arcsine square root transformation.[b] The percentage of each type of response (auditory, fusion, visual) is presented in Figure 2 (omissions and responses other than auditory, visual, and fusion were not taken into account[c]). Results were first analyzed in the VCL condition. For A/afa/V/aʃa/ (Fig. 2A), the distribution of responses across response types did not vary across groups (auditory based responses /afa/: CI: 16%, NH: 11%, fusion responses /asa/: CI: 6%, NH: 16%, visually based responses /aʃa/: CI: 72%, NH: 70%). A mixed ANOVA with response and group as factors revealed a main effect of response (F[1, 120] = 94.53, $p < 0.0001$). Planned comparisons with a Bonferroni correction showed that both groups of participants gave significantly more visually based responses than auditory based responses ($p < 0.0001$) or fusion responses ($p < 0.0001$). To analyze the effect of VR on the children's answers, we performed a mixed ANOVA with response type, visual condition (VCL and VR), and group as factors. It revealed main effects of response type (F[2, 120] = 33.38, $p < 0.0001$) and group (F[1, 60] = 6.11, $p = 0.02$) and a significant interaction between response type and visual condition (F[2, 120] = 82.28, $p < 0.0001$). Differences between each level of the visual condition factor were further investigated with separate ANOVAs for each response. Significant effects of visual condition were found: In both CI and NH groups, VR led to an increase of the number of auditory based responses (F[1, 61] = 118.49, $p < 0.0001$) and to a decrease of the number of visually based responses (F[1, 61] = 104.24, $p < 0.0001$). The number of fusions was not significantly modified ($p = 0.16$). The three-way interaction between group, visual condition, and response was not significant ($p = 0.86$), showing that this pattern of responses did not vary between groups, according to the visual condition (VCL or VR).

Contrary to the previous McGurk stimulus, the pattern of response to A/apa/V/aka/ in VCL varied between groups (Fig. 2B). The number of auditory based responses /apa/ (CI: 22%, NH: 53%), fusion /ata/ (CI: 36%, NH: 25%) and visually based responses /aka/ (CI: 38%, NH: 13%) varied across groups. A mixed ANOVA with response and group as factors confirmed these observations as it revealed a significant interaction between group and response (F[2, 120] = 10.72, $p < 0.0001$). Differences between groups were further investigated with separate ANOVAs for each level of the response factor. There was a significant group effect for the auditory based (F[1, 60] = 18.33, $p < 0.0001$) and the visually based (F[1, 60] = 10.93, $p < 0.005$) responses, showing that NH children gave significantly more auditory based responses and significantly less visually based responses than CI children. To analyze the effect of VR on the children's answers, we performed a mixed ANOVA with response, visual condition (VCL and VR), and group as factors. It revealed a main effect of response (F[2, 120] = 46.75, $p < 0.0001$), a significant interaction between response and group (F[2, 120] = 9.68, $p < 0.0001$) and between response, visual condition, and group (F[2,120] = 3.85, $p = 0.03$). Differences between each level of the visual condition factor were further investigated with separate ANOVAs for each response, with group as between-subject factor. First, concerning the "auditory based responses," there was a main effect of group (F[1, 60] = 11.97, $p < 0.005$), showing that NH children made more /apa/ answers than the CI group did. There was a main effect of visual condition (F[1, 60] = 128.03, $p < 0.0001$), showing that VR led to an increase in the number of auditory based responses. There was also a significant interaction between group and visual condition (F[1, 60] = 5.81, $p = 0.02$), showing that the increase of the auditory based responses was stronger in the CI group than in the NH group. Second, concerning the "fusion," there was a marginal main effect of group (F[1, 60] = 3.64, $p = 0.06$) indicating a tendency for CI children to make more fusion responses than the NH children could. There was also a main effect of visual

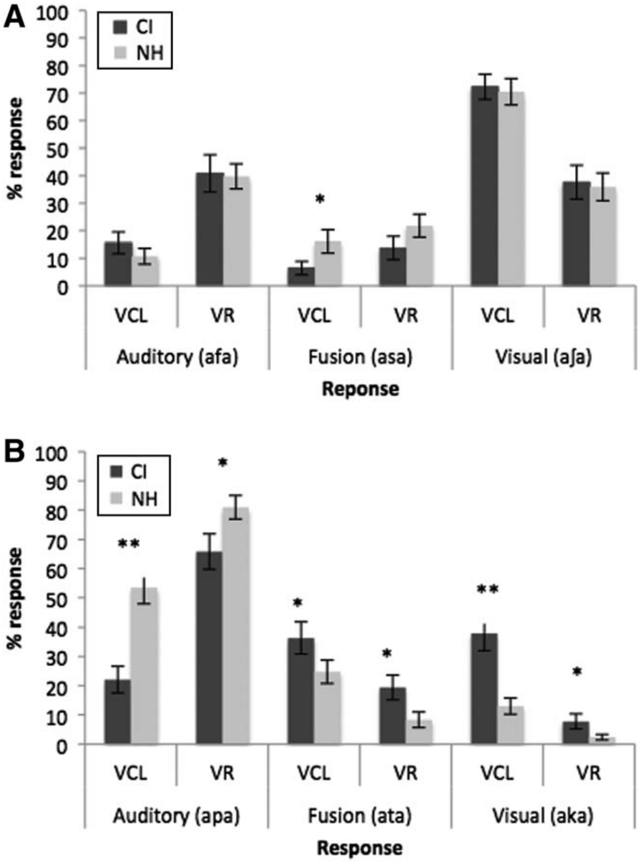

Fig. 2. A, Response type in the auditory/afa/ visual/aʃa/ McGurk trials in VCL blocks and in VR blocks for the CI children and the NH children (*$p < 0.05$; **$p < 0.001$). Bars represent the standard error of the mean. B, Response type in the auditory/apa/ visual /aka/ McGurk trials in VCL blocks and in VR blocks for the CI children and the NH children. VCL, visual clear; VR, visual reduction; CI, cochlear implanted; NH, normally hearing.

---

[b] The ANOVA and subsequent tests on mean scores yielded similar results.

[c] Responses other than auditory, visual, and fusion were mainly omissions. Only three CI children gave a few other responses (5 times altogether). Thereby, no test has been run on those other responses.

condition (F[1, 60] = 6.33, $p = 0.02$) indicating that VR led to a decrease in the number of fusions. This decrease was similar in both groups, because no significant interaction between visual condition and group was found ($p = 0.71$). Third, concerning the visually based responses, there was a main effect of group (F[1, 60] = 10.28, $p < 0.005$), showing that CI children made more of this response than NH children did. There was also a main effect of visual condition (F[1, 60] = 69.64, $p < 0.0001$), showing that VR led to a decrease in the number of visually based responses. The significant interaction between group and visual condition (F[1,60] = 6.44, $p = 0.001$) revealed that this decrease was stronger for CI group than for the NH group.

Impact of CI Proficiency

Results were next analyzed according to the auditory performance of each CI user. Participants were divided into groups of proficient (AO+; $n = 15$) and nonproficient (AO−; $n = 16$) CI users based on the median in the AO modality, averaged on VCL and VR conditions (79.17%). The AO+ and AO− groups had significantly different auditory performances (F[1, 30] = 35.59, $p < 0.001$). Among the seven children with a contralateral hearing aid, five belonged to the AO+ group and two belonged to the AO− group. As seen in Table 3, AO+ children and AO− children did not differ in terms of chronological age, age at implantation, and lipreading abilities (performance in the VO modality, averaged on VCL and VR conditions). In contrast, AO+ children had significantly higher performances in the AV modality (averaged on the VCL and VR conditions), (F[1, 30] = 8.61, $p = 0.006$) than the AO− children. A mixed ANOVA with visual condition and proficiency as factors showed that the AV gain was significantly decreased by VR, (F[1, 28] = 4.97, $p = 18.70$). There was no main effect of proficiency (F[1, 28] = 2.73, $p = 0.002$) nor significant interaction between proficiency and visual condition (F[1, 28] = 1.48, $p = 0.23$).

Responses to McGurk stimuli of both groups were compared in VCL and in VR conditions to investigate possible differences regarding the weighting of auditory and visual cues. The percentage of each type of answer to the McGurk stimuli (auditory, fusion, visual) is presented in Figure 3. For A/afa/V/aʃa/ (Fig. 3A), in VCL condition, a mixed ANOVA with response and proficiency as factors revealed a main effect of response (F[2, 58] = 61.72, $p < 0.0001$] and a significant interaction between proficiency and response (F[2, 58) = 3.67, $p = 0.03$). Differences between proficient and nonproficient CI children were further investigated with separate ANOVAs for each level of the response factor. Results showed that AO+ children made significantly more auditory based responses ($p = 0.03$) and significantly fewer visually based responses ($p = 0.04$) than AO− children. Next, when the visual condition factor was included in the ANOVA, analysis revealed a main effect of response (F[2, 58] = 18.74, $p < 0.0001$] and a significant interaction between response and visual condition (F[2, 58] = 38.69, $p < 0.0001$). Differences between each level of the visual condition factor were further investigated with separate ANOVAs for each response. In both AO− and AO+ groups, there was a significant effect for the auditory (F[1, 30] = 59.82, $p < 0.0001$) and visual (F[1, 30] = 54.26, $p < 0.0001$): auditory based responses were increased by VR whereas visually based responses were decreased. The three-way interaction Proficiency × Visual condition × Response was not significant ($p = 0.30$), indicating that the impact of VR on responses to McGurk stimuli was similar in AO+ and AO− children.

TABLE 3. Characteristics of the proficient (AO+) and the nonproficient (AO−) CI children

	AO+	AO−	Independent Sample *t* Tests
Age (mos)	127	122	ns
Age at implantation (mos)	42	43	ns
Mean correct identification in AO modality	89%	68%	$p < 0.001$
Mean correct identification in VO modality	51%	49%	ns
Mean correct identification in AV modality	94%	83%	$p < 0.01$
Mean AVG in VCL	61.39%	68.1%	ns
Mean AVG in VR	1.43%	26.62	ns

AO+, proficient CI users; AO−, nonproficient CI users; CI, cochlear implant; AV, audiovisual; ns, not significant; VO, visual-only; VCL, visual clear condition; VR, visual reduction condition.

For A/apa/V/aka/ (Fig. 3B), in VCL condition, a mixed ANOVA with response and proficiency as factors revealed no significant effects for either factor, showing that the distribution of responses did not vary across response types nor across

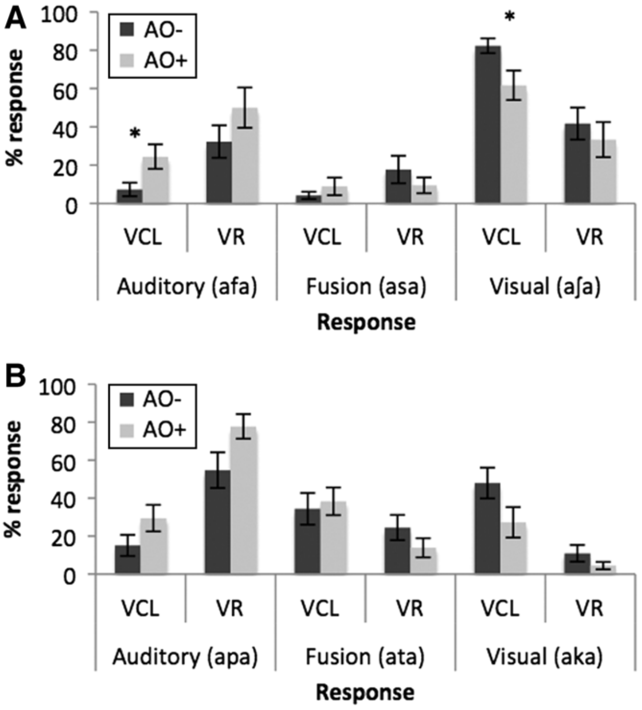

Fig. 3. A, Response type in the auditory/afa/ visual/aʃa/ McGurk trials in VCL blocks and in VR blocks for the CI children according to the CI proficiency. AO− represents the group of nonproficient CI users and AO+ represents the group of proficient CI users. B, Response type in the auditory/apa/ visual /aka/ McGurk trials in VCL blocks and in VR blocks for the CI according to the CI proficiency. AO−represents the group of nonproficient CI users and AO+ represents the group of proficient CI users. VCL, visual clear; VR, visual reduction; CI, cochlear implanted; NH, normally hearing.

TABLE 4. Characteristics of the proficient (NH+) and the nonproficient (NH−) NH children

	NH+	NH−	Independent Sample *t* Tests
Age (mos)	126	121	ns
Mean correct identification in AO modality	90%	75%	$p < 0.0001$
Mean correct identification in VO modality	58%	50%	ns
Mean correct identification in AV modality	93%	82%	$p < 0.0001$
Mean AVG in VCL	46.52%	27.45%	ns
Mean AVG in VR	1.49%	7.31%	ns

AV, audiovisual; NH$^+$, proficient NH children; NH–, nonproficient NH children; NH, normally hearing; ns, not significant; VO, visual-only; VCL, visual clear condition; VR, visual reduction condition.

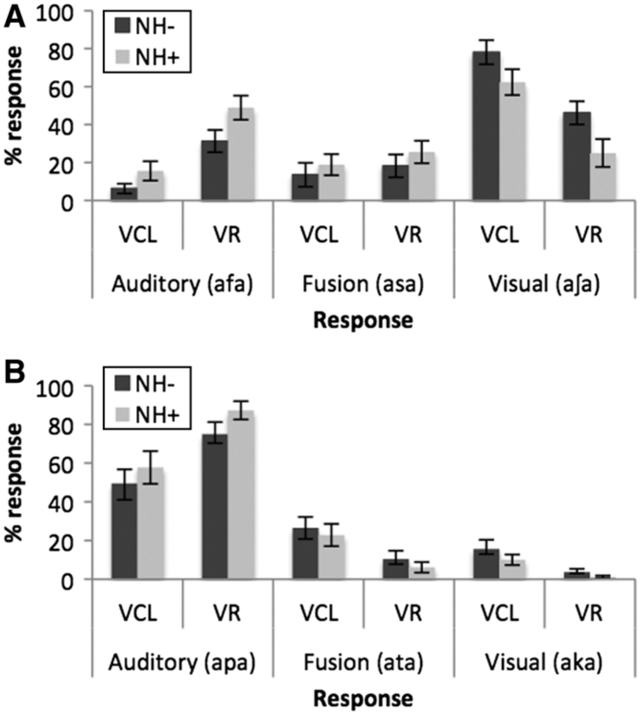

Fig. 4. A, Response type in the auditory/afa/ visual/aʃa/ McGurk trials in VCL blocks and in VR blocks for the NH children according to the auditory proficiency. NH− represents the group of nonproficient NH children and NH+ represents the group of proficient NH children. B, Response type in the auditory/apa/ visual /aka/ McGurk trials in VCL blocks and in VR blocks for for the NH children according to the auditory proficiency. NH− represents the group of nonproficient NH children and NH+ represents the group of proficient NH children. VCL, visual clear; VR, visual reduction; CI, cochlear implanted; NH, normal hearing.

groups. Next, when the visual condition factor was introduced in the ANOVA, a main effect of response (F[2, 58] = 18.74, $p < 0.0001$) and a marginal interaction between proficiency and response factors (F[2, 58] = 2.95, $p = 0.06$) were revealed. Further analysis revealed a significant effect of proficiency for auditory based responses (F[1, 30] = 4.01, $p = 0.04$) indicating that AO+ children made more /apa/ responses than AO− children. There was no significant effect of proficiency on the number of visually based responses ($p = 0.10$) nor on the number of fusions ($p = 0.94$). We also found a significant interaction between visual condition and response (F[2, 58] = 39.62, $p < 0.0001$). Further analysis revealed significant effects for each of the response factors: Auditory based responses were increased by visual reduction (F[1, 30] = 70.09, $p < 0.0001$), whereas fusion (F[1, 30] = 9.75, $p < 0.005$) and visually based responses (F[1, 30] = 44.62, $p < 0.0001$) were decreased. The three-way interaction between group, visual condition, and response was not significant ($p = 0.28$), showing that this pattern of responses according to the visual condition (VCL or VR) did not vary between groups.

Impact of Spectrally Reduced Speech Proficiency

Results of NH children were also analyzed according to their performance in the AO condition. Participants of the NH group were divided into proficient (NH+; $n = 15$) and nonproficient (NH−; $n = 16$) groups based on the median in the AO modality, averaged on VCL and VR conditions (84.72%). As seen in Table 4, NH+ and NH− children did not differ in terms of chronological age and lipreading abilities (VO modality). In contrast, NH+ children had significantly higher performance in AO (F[1, 30] = 49.19, $p < 0.0001$) and AV (F[1, 30] = 20.64, $p < 0.0001$) modalities than the NH− children. A mixed ANOVA with proficiency and visual condition as factors showed that AV gain was decreased significantly by VR (F[1, 30] = 6.33, $p = 0.02$). There was no main effect of proficiency (F[1, 30] = 0.006, $p = 0.94$) nor significant interaction between proficiency and visual condition (F[1, 30] = 1.55, $p = 0.11$).

The percentage for each type of answer (auditory, fusion, visual) is presented in Figure 4. For A/afa/V/aʃa/ (Fig. 4A), in VCL condition, a mixed ANOVA with proficiency and response as factors revealed a main effect of response (F[2, 58] = 40.95, $p < 0.0001$). A planned comparison with Bonferroni corrections showed that participants (NH+ and NH−) gave significantly more visually based responses than auditory based responses ($p < 0.0001$) and fusion ($p < 0.0001$). No difference between proficient and nonproficient NH children was found. Next, when the visual condition factor was introduced in the ANOVA, analysis revealed a main effect of response (F[2, 58] = 16.69, $p < 0.0001$), a significant interaction between response and proficiency (F[2, 58] = 3.81, $p = 0.03$) and between visual condition and response (F[2, 58] = 42.93, $p < 0.0001$). Further analysis revealed that over all visual conditions (VCL and VR), NH− children made significantly more visually based responses ($p = 0.02$) and significantly less auditory based responses ($p = 0.04$) than NH+ children did. Furthermore, analysis showed that VR led to a significant increase in the number of auditory based responses (F[1, 30] = 58.32, $p < 0.0001$) and to a significant decrease in the number of visually based responses (F[1, 30] = 48.66, $p < 0.0001$). The number of fusions were not modified by VR ($p = 0.09$). The three-way interaction for Proficiency × Visual Condition × Response was not significant ($p = 0.82$), indicating that the impact of VR on responses to McGurk stimuli was similar in both NH+ and NH− groups.

For A/apa/V/aka/ (Fig. 4B), in VCL, a mixed ANOVA with response and proficiency as factors revealed a main effect of response (F[2, 58] = 15.35, $p < 0.0001$). A planned comparison with Bonferroni corrections showed that participants (NH+ and NH−) gave significantly more auditory based responses than visually based responses ($p < 0.0001$) and fusion responses ($p = 0.01$). Next, when the visual condition was included in the ANOVA, analysis revealed a main effect of response (F[2, 58] = 67.21, $p < 0.0001$) and a significant interaction between response and visual condition factors (F[2, 58] = 41.53, $p < 0.0001$). Further analysis showed that auditory based responses were increased by VR (F[1, 30] = 57.68, $p < 0.0001$) whereas visually based responses (F[1, 30] = 24.88, $p < 0.0001$) and fusion (F[1, 30] = 29.63, $p < 0.0001$) were decreased. The three-way interaction between proficiency, visual condition, and response was not significant ($p = 0.72$), indicating that the impact of VR on responses to McGurk stimuli was the same in both groups.

In summary, in both groups, the visual-reduction technique had the following global impact:

(1) Performance in the AO condition increased.
(2) Performance in the VO condition decreased.
(3) The AV gain decreased.
(4) Regarding the McGurk stimuli, the number of visually based responses was always decreased whereas the number of auditory based responses always increased.

Modeling Results

Results from our study showed that VR led to an increase in the importance of audition in the response pattern. The question arises as to whether differences in AV performance are caused by differences in the informational content of auditory and visual modalities or by differences in the weight assigned to these modalities in the integration process itself (Grant 2002). To answer this question, a variant of the "fuzzy-logical model of perception" (FLMP), named "weighted fuzzy-logical model of perception" (WFLMP) was applied to our data. The standard form of the FLMP proposed by Massaro (1987, 1998) is a post-phonetic integration model with a statistically optimal integration rule. It can be expressed as follows:

$$P(R_i/A,V) = \frac{P(R_i/A)P(R_i/V)}{\sum_i P(R_i/A)P(R_i/V)}$$

In this expression, R_i and R_j are response categories, A and V are auditory and visual stimuli, $P(R_i/A)$, $P(R_i/V)$, and $P(R_i/A,V)$ are auditory (P_A), visual (P_V), and audiovisual (P_{AV}) response probabilities, respectively. Because the FLMP entails a fixed integration rule, a good fit of data to the FLMP means that any differences in AV responses are caused by differences in unisensory processing before the AV integration occurs.

For the study of subject variability, Schwartz (2010) used a variation of the FLMP, named the WFLMP, in which inputs from audition and vision are weighted. He compared this new model with the standard FLMP using various criteria (root-mean-square error, Bayesian model selection criterion) and found that WFLMP fitted the data better than the FLMP did. Therefore, WFLMP provides a meaningful indicator of how much individuals rely on audition and on vision (Schwartz et al. 2010). For these reasons, the WFLMP was used here. The WFLMP is defined by the following:

$$P(R_i/A,V) = \frac{P(R_i/A)^{\lambda_A} P(R_i/V)^{\lambda_V}}{\sum_j P(R_j/A)^{\lambda_A} P(R_j/A)^{\lambda_V}}$$

In this expression, λ_A and λ_V are subject-dependent factors used to weight the auditory and visual inputs in the computation of the AV responses. For each subject, a lambda value is defined between 0 and 1; λ_A and λ_V are computed from lambda by: $\lambda_A = \text{lambda}/(1-\text{lambda})$ and $\lambda_V = (1-\text{lambda})/\text{lambda}$, with thresholds maintaining λ_A and λ_V between 0 and 1. As a result, P_{AV} varies from a value close to P_A when lambda is close to 1, to a value close to P_V when lambda is close to 0, through a value identical to the FLMP prediction when lambda is close to 0.5, with λ_A and λ_V both equal to 1 (Schwartz 2010).

The assessment criteria used here was the root-mean-square error (RMSE), computed by taking the squared distance between observed and predicted probabilities of responses, averaging them over all categories C_i and all experimental conditions E_j (naming AO, VO, and AV) and taking the square root of the result.

$$\text{RMSE} = \left[\frac{\left(\sum_{E_j,C_i}\left(P_{E_j}(C_i) - P_{E_j}(C_i)\right)^2\right)}{n_E n_C}\right]^{1/2}$$

In this equation, observed probabilities are in lower case and predicted probabilities are in upper case. Nonresponse data are included as a supplementary output category to sum at 1 for each input category.

Figure 5 shows that regression lines of both CI and NH groups were flat in the VCL condition and steeper in the VR condition where RMSE decreases when lambda increases. This result suggests that the model fits the data in VR condition better when an additional weight is applied to the auditory modality. To test this hypothesis, we performed an analysis of covariance with RMSE as the dependent variable, visual condition (VCL, VR), and group as factors and lambda as covariate factor. We found a significant effect of visual condition (F[1, 75] = 49.54, $p < 0.0001$), a significant effect of lambda (F[1, 75] = 81.11, $p < 0.0001$), and a significant interaction between lambda and visual condition (F[1, 75] = 9.82, $p = 0.003$). This significant interaction indicates that the variation of RMSE according to lambda was not the same in VCL condition than in VR condition. There was no main effect group ($p = 0.07$) and no significant interaction involving the group variable. Taken together, results indicate that the weight of audition is increased in VR condition compared with VCL condition. There is no evidence that this increased auditory weight varies across groups.

DISCUSSION

Perception of incongruent AV stimuli by CI children is largely biased toward the visual modality (Schorr et al. 2005, Leybaert & Colin 2007). Here, we examined whether this visual bias is fixed and immutable by varying the quality of the visual speech cue and the degree of auditory proficiency. To that aim, a group of CI children was presented with an AV speech-perception paradigm in which stimuli were either clear or visually

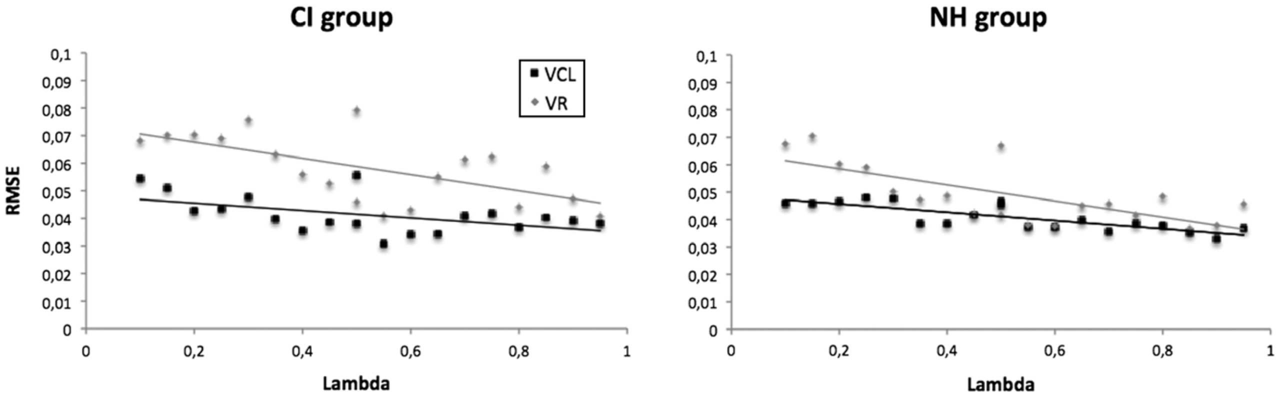

Fig. 5. Variation of the root-mean-square error as a function of the lambda parameter tuning fusion in the weighted fuzzy-logical model of perception. CI, cochlear implanted; NH, normally hearing.

degraded. Their performance was compared with those of NH children receiving spectrally reduced speech. Results of both groups were similar in AO, AV, and VO modalities, indicating comparable levels of auditory performance, lipreading, and AV gain. Results of both groups confirmed the effectiveness of our visual-reduction technique, because performance was decreased in the VO and AV modalities. On the contrary, performance in the AO modality was increased. This was an unexpected effect because no visual cue was provided in that modality (the speaker's face remained still). One explanation might be that unreliable visual information in the surrounding AV and VO trials led to an increase of attention given to the auditory information.

Responses to McGurk stimuli were examined separately. In the case of A/afa/V/aʃa/ presented in VCL, CI children gave mainly visually based responses /aʃa/. This result corroborates previous findings (Schorr et al. 2005; Leybaert & Colin 2007) and shows that when presented with incongruent AV stimuli, CI children base their perception mainly on the visual modality. NH children also choose the visual modality, contrary to previous findings showing a tendency of NH children to choose the auditory modality (McGurk & MacDonald 1976; Massaro 1984; Massaro et al. 1986; Tremblay et al. 2007). Here, because of the spectrally reduced speech, NH children had to base their perception on vision, which had become their most reliable source of information. This hypothesis is confirmed by looking at the confusion matrices (Appendix A1, Supplemental Digital Content 1, http://links.lww.com/EANDH/A87): visual /aʃa/ is much better identified than auditory /afa/ (90% versus 50% correct).

Next, when the visual input was reduced, the number of visually based responses was decreased in both groups, whereas the number of auditory based responses was increased. Again, we might say that responses of CI and NH children are dependent on the informational content of both unimodal channels. In VR condition, the better identification of auditory /afa/ compared with visual /aʃa/ (see Appendix A1, Supplemental Digital Content 1, http://links.lww.com/EANDH/A87) explains the significant increase of auditory based responses in both CI and NH children. However, the informational content of each input does not seem to be the only explanation for the increase in the number of auditory based responses. Indeed, analyses made in the framework of the WFLMP showed that the weight of audition was increased in both groups when the visual input was degraded. This important result indicates that our technique of VR modified the informational content of the VO modality and that consequently, participants awarded more weight to audition. Children in this study acted as if they were forced to lean more on audition because of the VR. In sum, differences between VCL and VR were not only caused by differences in unimodal performance but also by differences in AV speech integration per se.

In the case of A/apa/V/aka/ presented in VCL, CI children gave mainly visually based responses and fusion. The fact that the fusion rate was high might come as a surprise because previous studies have shown weak fusion capacities in CI children (Schorr et al. 2005; Leybaert & Colin 2007). Note, however, that some of the errors made by the CI group in AO and VO conditions were consistent with the fusion responses (participants answered /ata/ on 18% of the trials in response to /apa/ in AO modality and on 9% of the trials in response to /aka/ in VO modality). This, as argued by Grant & Seitz (1998), makes it difficult to distinguish between auditory/visual errors (with no AV integration) and evidence for McGurk fusion (with AV integration). Accordingly, the fusion rate /ata/ in the CI group might be partly explained by unimodal errors. Conversely, NH children gave mainly auditory based responses. This is consistent with previous findings (McGurk & MacDonald 1976; Massaro, 1984; Massaro et al. 1986; Tremblay et al. 2007). Thus, in VCL, CI children and NH responded differently. CI children leaned on the visual modality whereas NH children leaned on the auditory modality. This difference can be explained by differences in unimodal performance. In NH children, auditory /apa/ is better identified (93%) than visual /aka/ (59%) is. Thereby, NH children base their perception on audition. In CI children, auditory /apa/ (61%) and visual /aka/ (66%) are similarly identified. The fact that CI children leaned more on vision than audition shows, once again, that vision is the most reliable source of information for children with a CI. Next, when vision was degraded, the number of visually based responses was decreased in both groups, whereas the number of auditory based responses was increased. Again, this result relates to the informational content of AO and VO modalities. However, the improved identification of auditory /apa/ compared with visual /aka/ in VR condition (see confusion matrices, Appendix A2, Supplemental Digital Content 2, http://links.lww.com/EANDH/A86) only partly explains the significant increase of auditory based responses in both CI and NH children. The other explanation is that, as for the previous McGurk stimulus, participants of both groups awarded more weight to the auditory modality.

In sum, VR led to a major increase of the weight of audition in the pattern of response. This increase is related not only to the degree of informational content of each unimodal channel but also to the fact the weight of audition, in the AV speech integration process, was increased.

A further result relates to the impact of VR on fusion responses. In both groups, the fusion rate was decreased in the case A/apa/V/aka/ (fusion /ata/) and unchanged in case of A/afa/V/aʃa/ (fusion /asa/). The decrease in the number of /ata/ fusion is consistent with a decrease in the number of McGurk illusions when vision is degraded, as found in previous studies (Fixmer & Hawkins 1998; MacDonald et al. 2000). The fact that the number of /asa/ fusions did not decrease can be explained in the NH group by the increase in the number of confusions made in the AO modality. Indeed, confusion matrices show that NH children answered /asa/ in response to /afa/ in 15% of the trials in the AO VR modality, whereas this confusion never happened in VCL. Thereby, it is possible that the number of real fusions (with integration) was decreased by VR but this decrease was masked by the emergence of confusions (without integration). In the CI group, the number of fusions was already very low in VCL, possibly precluding any further decrease.

Another issue addressed in our study concerned the impact of auditory proficiency. Regarding the CI group, our data showed that when the visual input was clear, all CI children (proficient and nonproficient) gave mainly visually based responses to McGurk stimuli. This was especially true for A/afa/V/aʃa/, where the visual cue /aʃa/ is more salient than the auditory cue /afa/ (80% correct versus 44% and 89% correct *vs*. 78% for the AO- and AO+ respectively). For A/apa/V/aka/, the visual preference was less obvious and responses were well balanced between the three possibilities. This result can be explained by the fact that auditory /apa/ and visual /aka/ are similarly perceived in the unimodal conditions (61% versus 70% in the AO- group and 71% versus 61% in the AO+ group).

Nevertheless, the nonproficient CI users seemed to put more weight on vision than the proficient CI users because their rate of visually based responses was significantly higher (in both McGurk stimuli), even though their lipreading performances were similar. In addition, they paid less attention to audition than the proficient users did because their level of auditory based responses was significantly lower (also for both McGurk stimuli). These data suggest a relationship between AV speech perception and CI proficiency. Such a link was also reported in a recent study (Tremblay et al. 2010) with postlinguistically deaf adults. In their study, no significant difference in the number of McGurk fusions was found between proficient and nonproficient CI users and a group of NH controls. However, group differences emerged in the response alternatives chosen. When a nonfused response was produced, NH controls and proficient CI users tended to choose an auditory alternative whereas nonproficient CI users leaned more toward visual alternatives. These data suggest that, for postlinguistically deaf adults (contrary to prelinguistically deaf children), the AV imbalance in favor of vision is only present in CI users whose auditory input is degraded because of poor CI proficiency. Another study examining the impact of CI proficiency (Champoux et al. 2009) found that the recognition of auditory words was impaired in nonproficient CI users when irrelevant visual stimuli (dot motion and lip motion) were simultaneously presented. As this was not the case for proficient CI users, this result confirms that CI proficiency plays an important role in the ability to segregate vision from audition.

However, in our study, even the proficient CI children gave mainly visually based responses to McGurk stimuli. This differs from postlinguistically deaf adults and is probably because of the fact that prelinguistically deaf children were never confronted with AV associations until the day their CI was switched on. Until that moment, they had to rely only on their lipreading skills to access speech recognition. At that point, in a situation of conflicting AV cues, they continued to use this strategy and relied on the least ambiguous modality.

The impact of auditory proficiency was also examined in NH children. We found no significant difference between proficient and nonproficient NH children when the visual input was clear: for A/afa/V/aʃa/, both groups relied on vision, probably because of a better identification of V/aʃa/ than A/afa/ in the unimodal conditions (92% versus 45% for the NH− and 89% versus 54% for the NH+). For A/apa/V/aka/, both groups relied on audition, probably because of a better identification of A/apa/ than V/aka/ in the unimodal conditions (92% versus 57% for NH− and 96% versus 61% for NH+). Some differences appeared when the visual input was degraded, at least for A/afa/V/aʃa/: NH+ children made significantly more auditory based responses and significantly less visually based responses than NH− children. Once again, this result relates to performance in unimodal conditions: A/afa/ was correctly identified at a rate of 63% for the NH− group and at a rate of 82% for the NH+ group. On the contrary, because auditory /apa/ was identified at a same rate in both groups, no group difference appeared in VR condition for A/apa/V/aka/. To conclude, differences in responses to McGurk stimuli according to spectrally reduced signal proficiency are directly related to performance in AO modality.

CONCLUSIONS

Our study confirms the existence of a natural bias in favor of vision in AV speech perception in CI children when audition and vision are incongruent (in VCL condition). Nevertheless, this visual bias strongly depends on the experimental situation, because degrading the visual cue leads to an increase of the weight of audition in the response pattern. The question arises as to whether differences between visual clear and VR were caused by differences in integration per se or rather by differences in the informational content of each sensory channel (Grant 2002). To answer that question, the data set was analyzed in the framework of the FLMP (Massaro 1987, 1998). More precisely, we used the variant of the FLMP, the WFLMP, as proposed by Schwartz (2010). We found that the model fitted the data better when an additional weight was applied to the auditory modality. This was true for both groups. In other words, differences between visual-clear cues and visual-reduced cues were not only caused by difference in the informational content of the VO modality, but also by a supplementary auditory weight. It suggests that AV speech integration, in CI children and in NH children, is actually more an adaptive process than has been presented in previous research (Schorr et al. 2005; Leybaert & Colin 2007; Rouger et al. 2008). Indeed, results of both groups demonstrate that, in AV speech perception, responses are influenced by the task properties, here VR (supporting the view of Schwartz, 2010). This was also true in our previous study (Huyse et al. in revision) in which not only VR but also auditory noise influenced the AV speech perception. In VCL, a stationary noise generated mainly visually based responses whereas a modulated noise allowed an increase of the number of McGurk fusions. Conversely, VR led to an increase in auditory based responses, whether the noise was stationary or modulated. Thus, the importance of audition and vision in speech perception depends on the experimental situation. The experimental situation modulates the level of certainty the perceiver has regarding auditory and visual information and the weight awarded to each modality. CI and NH children gave more importance to the modality that was more informative. This suggests that CI children are not trapped in a situation where they will only rely on vision. Rather, they will use the more reliable modality, whether auditory or visual. It is interesting that the more reliable modality was not always the same for CI children than for NH children (e.g., the visual modality was more informative than the auditory modality in the case of A/apa/V/aka/), suggesting that CI children should not be considered NH children. Moreover, our study showed that AV speech perception also depends on CI proficiency, with proficient CI children being "less visual" and "more auditory" than the nonproficient, also because of differences in modality informational content. However, even for proficient CI users, lipreading remains an essential source of information. Indeed, the CI is highly sensitive to noise (Fu et al. 1998; Fetterman & Domico 2002; Munson & Nelson 2005), leading to decreased auditory performance in many common situations (classrooms, restaurants, etc.). Because auditory and visual cues are congruent in everyday life, lipreading helps CI users to overcome the difficulties, and it should therefore be part of the rehabilitation programs for CI patients.

ACKNOWLEDGMENT

The study was funded by a grant from the Fonds National de Recherche Scientifique (FNRS), Belgium (Aurélie Huyse is aspirante FNRS) and by Fonds de la Recherche Fondamentale Collective grant 2.4621.07 (Seeing to Hear Better: Normal and Atypical Development).

The authors have no conflicts of interest to disclose.

Address for correspondence: Aurélie Huyse, LCLD, 50 Avenue F.D. Roosevelt—1050 Brussels, Belgium. E-mail: ahuyse@ulb.ac.be

Received March 15, 2011; accepted June 20, 2012.

REFERENCES

Bergeson, T. R., Pisoni, D. B., Davis, R. A. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear Hear*, *26*, 149–164.

Berthommier, F. (2001). Audio-visual recognition of spectrally reduced speech. Dans AVSP 2001-International Conference on Auditory-Visual Speech Processing Aalborg, Denmark.

Champoux, F., Lepore, F., Gagné, J. P., et al. (2009). Visual stimuli can impair auditory processing in cochlear implant users. *Neuropsychologia*, *47*, 17–22.

Desai, S., Stickney, G., Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *J Acoust Soc Am*, *123*, 428–440.

Dowell, R. C., Martin, L. F., Tong, Y. C., et al. (1982). A 12-consonant confusion study on a multiple-channel cochlear implant patient. *J Speech Hear Res*, *25*, 509–516.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J Speech Hear Res*, *12*, 423–425.

Fetterman, B. L., & Domico, E. H. (2002). Speech recognition in background noise of cochlear implant patients. *Otolaryngol Head Neck Surg*, *126*, 257–263.

Fixmer, E., & Hawkins, S. (1998). The influence of the quality of information on the McGurk effect. Paper presented at the international conference on audio-visual speech processing, Terrigal, Australia December 4–6, 1998.

Fu, Q. J., Shannon, R. V., Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *J Acoust Soc Am*, *104*, 3586–3596.

Füllgrabe, C., Berthommier, F., Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hear Res*, *211*, 74–84.

Geers, A. E., Nicholas, J. G., Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear Hear*, *24*(1 Suppl), 46S–58S.

Giraud, A. L., Price, C. J., Graham, J. M., et al. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron*, *30*, 657–663.

Grant, K. W. (2002). Measures of auditory-visual integration for speech understanding: A theoretical perspective. *J Acoust Soc Am*, *112*, 30–33.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *J Acoust Soc Am*, *104*, 2438–2450.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am*, *108*(3 Pt 1), 1197–1208.

Kaiser, A. R., Kirk, K. I., Lachs, L., et al. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J Speech Lang Hear Res*, *46*, 390–404.

Kiefer, J., Hohl, S., Stürzebecher, E., et al. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology*, *40*, 32–42.

Lachs, L., Pisoni, D. B., Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear Hear*, *22*, 236–251.

Leybaert, J., & Colin, C. (2007). Le rôle des informations visuelles dans le développement du langage de l'enfant sourd muni d'un implant aire [The role of visual information in the language development of deaf children with cochlear implants]. *Enfance*, *59*, 245–253.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

MacDonald, J., Andersen, S., Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception*, *29*, 1155–1168.

MacLeod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*, 21(2), 131–141.

Manuel, S. Y., Repp, B. H., Studdert K. M., Liberman, A. M. (1983). Exploring the "McGurk effect." *J Acoust Soc Am*, *74*, S66.

Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press.

Massaro, D. W., Thompson, L. A., Barron, B., Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *J Exp Child Psychol*, *41*, 93–113.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Mohammed, T., Campbell, R., MacSweeney, M., et al. (2005). Speechreading skill and visual movement sensitivity are related in deaf speechreaders. *Perception*, *34*, 205–216.

Moody-Antonio, S., Takayanagi, S., Masuda, A., et al. (2005). Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otol Neurotol*, *26*, 649–654.

Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *J Acoust Soc Am*, *118*, 2607–2617.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *J Exp Psychol Hum Percept Perform*, *22*, 318–331.

Rouger, J., Fraysse, B., Deguine, O., et al. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Res*, *1188*, 87–99.

Rouger, J., Lagleyre, S., Fraysse, B., et al. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci USA*, *104*, 7295–7300.

Schorr, E. A., Fox, N. A., van Wassenhove, V., et al. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proc Natl Acad Sci USA*, *102*, 18748–18750.

Schwartz, J. L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *J Acoust Soc Am*, *127*, 1584–1594.

Schwartz, J. L., Tiippana, K., Andersen, T. (2010). Disentangling unisensory from fusion effects in the attentional modulation of McGurk effects: A Bayesian modeling study suggests that fusion is attention-dependent. *AVSP*, Japan 23–27.

Sekiyama, K., Burnham, D., Tam, H., Erdener, D. (2003). Auditory-visual speech perception development in Japanese and English speakers. AVSP 2003-International Conference on Audio-Visual Speech Processing, ISCA September 4–7, France.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *J Phon*, *21*, 427–444.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Skinner, M. W., Fourakis, M. S., Holden, T. A., et al. (1999). Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies II. Consonants. *Ear Hear*, *20*, 443–460.

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 Hear*, *26*, 263–275.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J Acoust Soc Am*, *26*, 212–215.

Tremblay, C., Champoux, F., Lepore, F., et al. (2010). Audiovisual fusion and cochlear implant proficiency. *Restor Neurol Neurosci*, *28*, 283–291.

Tremblay, C., Champoux, F., Voss, P., et al. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE*, *2*, e742.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., et al. (1997). Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *J Acoust Soc Am*, *102*, 508–522.

Walden, B. E., Prosek, R. A., Montgomery, A. A., et al. (1977). Effects of training on the visual recognition of consonants. *J Speech Hear Res*, *20*, 130–145.

Zeng, F. G. (2004). Trends in cochlear implants. *Trends Amplif*, *8*, 1–34.

REFERENCE NOTE

1. Grosgeorge, A. (2005). Etudes des caractéristiques perceptives de la parole réduite chez les sujets sains et les implantés cochléaires [Studies of perceptual characteristics of speech reduced in healthy subjects and cochlear implants]. Unpublished doctoral dissertation, Institut de la Communication Parlée, Grenoble, France.