# Evidence that cochlear-implanted deaf patients are better multisensory integrators

J. Rouger*, S. Lagleyre†, B. Fraysse†, S. Deneve‡, O. Deguine*†, and P. Barone*§

*Centre de Recherche Cerveau et Cognition, Université Toulouse 3, Centre National de la Recherche Scientifique, Faculté de Médecine de Rangueil, 31062 Toulouse Cedex 9, France; †Service d'Oto–Rhino–Laryngologie, Hopital Purpan, FR-31059 Toulouse, France; and ‡Groupe de Neurosciences Théoriques, École Normale Supérieure, 3, Rue d'Ulm, 75005 Paris, France

Edited by Eric I. Knudsen, Stanford University School of Medicine, Stanford, CA, and approved February 26, 2007 (received for review October 24, 2006)

**The cochlear implant (CI) is a neuroprosthesis that allows profoundly deaf patients to recover speech intelligibility. This recovery goes through long-term adaptative processes to build coherent percepts from the coarse information delivered by the implant. Here we analyzed the longitudinal postimplantation evolution of word recognition in a large sample of CI users in unisensory (visual or auditory) and bisensory (visuoauditory) conditions. We found that, despite considerable recovery of auditory performance during the first year postimplantation, CI patients maintain a much higher level of word recognition in speechreading conditions compared with normally hearing subjects, even several years after implantation. Consequently, we show that CI users present higher visuoauditory performance when compared with normally hearing subjects with similar auditory stimuli. This better performance is not only due to greater speechreading performance, but, most importantly, also due to a greater capacity to integrate visual input with the distorted speech signal. Our results suggest that these behavioral changes in CI users might be mediated by a reorganization of the cortical network involved in speech recognition that favors a more specific involvement of visual areas. Furthermore, they provide crucial indications to guide the rehabilitation of CI patients by using visually oriented therapeutic strategies.**

cochlear implant | deafness | multisensory integration | speech comprehension

Despite the apparent division between sensory modalities from the receptors to high cortical levels, we can simultaneously integrate visual and auditory signals resulting in qualitative percepts distinct from those derived from a single unisensory stimulus (1, 2). Furthermore, in cases of precise temporal or spatial congruency between the bisensory stimuli, multisensory integration is expressed at the behavioral level by perceptual improvements by reducing ambiguity (3, 4) and at the neuronal level by enhancing neuronal activity (5). Multisensory integration is also essential for speech recognition, which is based on the simultaneous integration of visual information derived from lip movements and auditory cues produced by the talker (6). The McGurk effect, in which a mismatch between the visual and auditory speech signals is artificially introduced, reveals that the visual information derived from lip movements can strongly influence our auditory perception (7). Although we might not be aware of the relevance of the visual cues for normal speech recognition, the influence of vision becomes convincingly apparent when the auditory information is embedded in noise. In degraded auditory conditions, the visuoauditory presentation leads to higher performance of recognition, when compared with the auditory alone stimulation (8, 9), in a mechanism that mimics an improvement in the acoustic signal-to-noise ratio (SNR) (10).

In normally hearing (NH) subjects, although speechreading performance is very low, the association during development between the auditory and visual speech information is critical for a normal acquisition of multisensory speech perception (11). Speechreading can become extremely crucial in the case of profound deafness because the acquisition of strong skills in speechreading is one of the sensory substitution strategies developed by deaf patients to access speech recognition. During recent decades, remarkable technical progress has been made in the efficiency of cochlear implants (CIs), allowing patients afflicted with sensorineural hearing loss to recover a large range of auditory functions (12–14). Although in postlinguistically deafened adults the implantation of a cochlear neuroprosthesis allows a significant recovery of auditory speech intelligibility (13, 15), CI patients remain highly sensitive to noisy environments and have impaired speech-recognition performance in the presence of masking sounds (16). This suggests that CI users might develop specific visual and visuoauditory integration skills to overcome their difficulties in speech recognition in everyday life. Although it is known that deaf people develop specific visual abilities (17, 18), including speechreading (19), no data are available on the postimplantation evolution of visual performance in CI users and how these abilities are related to the recovery of the auditory functions. To answer these questions, we have analyzed the longitudinal postimplantation performance of CI patients in word recognition presented in a uni- or bisensory modality. For further analysis, we compared CI users' performance to that of NH subjects who were presented with identical stimuli, with an additional sound computation designed to simulate the processing performed by a CI system. These behavioral data, combined with a multisensory integration model, reveal that, as a consequence of their auditory deprivation, CI deaf patients have developed a much higher proficiency to fuse visual speech information to auditory cues as compared with NH subjects.

## Results

**Pre- and Postimplantation Performance of Speech Recognition in CI Patients.** We have analyzed the performance of 97 CI users in disyllabic word recognition by using three modalities (auditory, visual, and visuoauditory) during a longitudinal study that extended over 8 years after implantation (Fig. 1).

**Auditory Speech.** First, at the time the implant is switched on (T0), CI users obtain a significant recovery of word recognition in auditory modality, with a performance level of 47.1 ± 27.3% SD in quiet conditions. This performance level is much higher than that obtained before implantation by using an external hearing aid (mean 10.4 ± 14.2% correct, $P < 0.05$). Auditory performance increases significantly during the subsequent months ($P < 0.05$), before reaching a plateau from about the seventh month on and then showing no significant improvement in the following years (mean 81% over the first year).

Author contributions: J.R. and S.L. contributed equally to this work; J.R., S.L., B.F., S.D., O.D., and P.B. designed research; J.R., S.L., B.F., O.D., and P.B. performed research; J.R., S.L., S.D., and P.B. analyzed data; and J.R., S.L., S.D., O.D., and P.B. wrote the paper.

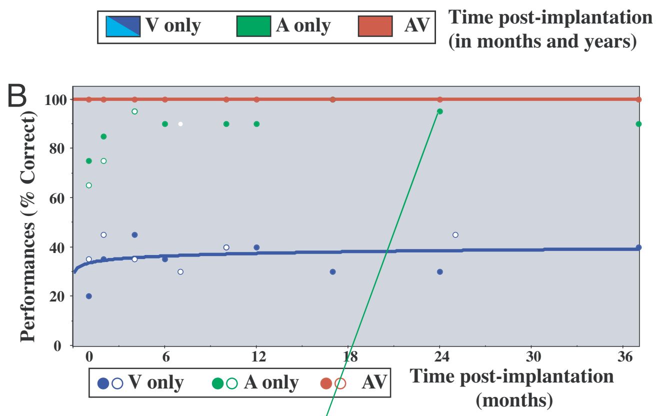
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: CI, cochlear implant; NH, normally hearing; SNR, signal-to-noise ratio.

See Commentary on page 6883.

§To whom correspondence should be addressed. E-mail: pascal.barone@cerco.ups-tlse.fr.




**Speechreading.** At T0, speechreading performance in CI users is elevated and impressively higher than that observed for NH subjects tested with the same talker (35.1 ± 14.7% vs. 9.4 ± 7.1%, respectively; $P < 0.05$). This speechreading ability in CI users is similar to that obtained a few months before implantation (mean 30.1 ± 15.1%; paired test, $P = 0.62$) and remains unchanged across all postimplantation periods tested (>35%; $P > 0.05$), although CI users have reached their maximal auditory performance. Furthermore, visual performance at T0 is not correlated with auditory proficiency ($r^2 = 0.001$, $P = 0.76$). At the time of implantation, duration of deafness is not correlated to visual performance ($r^2 = 0.001$, $P = 0.77$). This latter result should be tempered because most of the patients were suffering from a progressive hearing impairment, such that deafness duration could hardly be reliably defined. However, in three CI users who were affected by sudden deafness (such as meningitis) and implanted only 1 year later, speechreading performance levels were similar to that of the CI population (20%, 30%, and 45%, respectively). To strengthen this observation, we have included in our analysis data obtained from five supplementary CI users (not included in the longitudinal retrospective study) suffering from sudden deafness occurring within <1 year of implantation. In this enlarged sample ($n = 8$), we observed that performance in visual-only conditions was much higher than that observed in NH subjects (mean 27.5 ± 10.7% vs. 9.4%, respectively; $P < 0.05$). In these CI users, several months or years of auditory recovery postimplantation (auditory-only performance >90% correct) did not affect their speechreading performance. Despite the limited number of observations, this suggests that a high level of speechreading ability can be acquired rapidly during a period of auditory deprivation and then remain at stable values.

**Audiovisual Speech.** As expected from the classically perceptual benefit derived from multisensory integration (5), prior to implantation, CI users present higher performance in visuoauditory conditions compared with the auditory-alone conditions (55.8 ± 21.0% vs. 10.4%, respectively; $P < 0.05$). A similar effect is observed in CI users postimplantation; when compared with unisensory conditions, visuoauditory integration results in an improvement in word recognition in CI patients tested at T0 (86 ± 17.4% correct; $P < 0.05$ for both comparisons). From the time of implantation, audiovisual recognition improved slightly ($P < 0.05$) with practice, allowing CI users to reach near-perfect performance levels (94 ± 12.0%) as early as the second month postimplantation.

We believe that the difference in bisensory performance of CI users when comparing pre- and postimplantation periods (55.8% vs. 86% at T0; $P < 0.05$) is derived from higher auditory performance provided by the neuroprosthesis. In agreement with this, we saw that, in a limited number of CI users ($n = 14$) who did not show any improvement in auditory word recognition, the visuoauditory gain remains unchanged when comparing the two testing periods (mean visuoauditory benefit preimplantation 0.54 vs. 0.62 at T0; paired test, $P = 0.23$).

**Comparison of Performance of NH and CI Subjects.** Our results show that, during the period of deafness, CI patients have developed a specific ability in speechreading that distinguishes them from the poor speechreading skills of NH listeners. We hypothesized that this high visual aptitude might induce in CI users an improvement of the mechanisms of multisensory integration, leading to greater visuoauditory benefits than those observed in NH subjects. In this scheme, we compared the visuoauditory gain in CI users at T0 (i.e., without training) to the one obtained in naïve NH subjects exposed to a degraded auditory signal. This auditory degradation allows us to make direct comparisons of visuoauditory performance from both groups at equivalent ranges of nonoptimal auditory performance. To degrade the auditory performance of NH subjects, we first used a masking paradigm with white noise at different SNRs. In these protocols, we observed a higher recognition rate in visuoauditory versus auditory-only conditions (Fig. 2*A*), especially at intermediate SNRs at ≈15 dB (20). Second, we used a noise-band vocoder paradigm with different frequency bands that simulates the processing strategy of CIs (21). In this simulation, the global temporal and spectral information of the signal are preserved, whereas the fine temporal cues within each spectral component are removed. In this case, performance (auditory and visuoauditory) decreases rapidly as the number of bands decreases, leading to near-zero values in the two-electrode simulation (1.5% mean recognition in auditory-only presentation). However, whereas bisensory presentation improved NH subjects' performance (Fig. 2*A*), the visuoauditory gain was much lower than the one obtained in the masking protocol at equivalent auditory performance levels, suggesting that visuoauditory integration mechanisms of speech perception strongly depend on the integration of fine spectrotemporal auditory information. This hypothesis was confirmed by our model (see *Are CI Patients Better Multisensory Integrators*). When visuoauditory performance of CI users is compared with that of the NHs exposed to degraded auditory stimuli, we show that the visuoauditory gain in CI patients is higher than that observed in NH subjects in the simulation or noise-masked conditions (both comparisons, $P < 0.001$; Fig. 2*B*). The differences in favor of CI users are especially

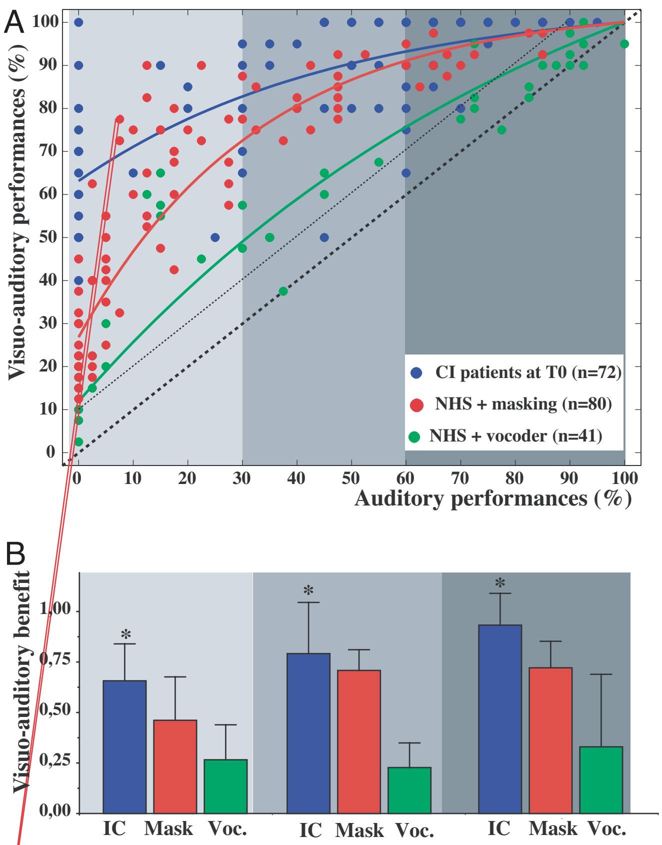



**Fig. 2.** Relationships between auditory and visuoauditory performance and bisensory gain. (*A*) For each group we have plotted the performance of individual subjects in auditory-only conditions with respect to the performance in visuoauditory conditions. Each point corresponds to a single subject tested in a single condition. This graph shows that the three experimental groups (CI users at T0, NH subjects with masking noise, and NH subjects with vocoder simulations) are clearly segregated in the correlation graph, indicating that the three populations present different gains after a bisensory presentation. (*B*) The visuoauditory gain (expressed as [(VA-A)/(100-A)]) is presented separately for three ranges of auditory performance (<30% correct, between 30% and 60% correct, and >60% correct) from each group. This subdivision shows that, in all cases, CI patients obtained a significantly higher visuoauditory gain than that observed in NH subjects tested with degraded auditory stimuli. The visuoauditory benefit for CI users is considerably greater than that computed from the NH subject vocoder group. The gains from these latter groups are directly comparable because of a similar processing of the auditory stimulus. Asterisks indicate statistically significant differences in visuoauditory benefits between CI users and both NH subject groups.

large in conditions of low auditory performance, where the range of correct recognition falls to <30% (CIPs vs. NHs, $P < 0.01$ for both comparisons). For example, a subset of CI users ($n = 13$) unable to perform auditory identification at all (0% correct) showed a high level of performance in the visuoauditory condition (mean 63% correct). In contrast and compared with this subset of CI users, NH subjects showing a similar level of auditory word recognition (0% correct, $n = 19$) due to highly degraded auditory conditions never reached the visuoauditory performance levels (mean 25.4% and 12.5% in masking or vocoder simulation protocols, respectively).

Although in CI patients the high efficiency of bisensory word recognition was not correlated to the level of speechreading ($r^2 = 0.068$, $P < 0.05$; see ref. 22), we further tested the hypothesis that the difference between CI and NH bisensory integration could be due to differences in absolute levels of visual performance. Consequently, we selected a subgroup of CI users showing low visual performance (lower than 20%; $n = 15$). We found within this group that the visuoauditory gain was still higher than that of NHs engaged in the simulating protocol (0.52 vs. 0.26, $P < 0.001$). In our opinion, this reinforces our conclusion that CI users have acquired a higher bisensory proficiency per se compared with NH subjects.

**Are CI Patients Better Multisensory Integrators?** As mentioned previously, the better performance levels of CI users compared with NHs with simulated implants could be due either to their stronger visual performance or to a better capacity for integrating visuoauditory inputs. Furthermore, electrophysiological studies have challenged whether the rules governing neuronal computing during multisensory interactions are superadditive, additive, or subadditive (23). Does it apply to the performance of speech recognition in bisensory conditions? To evaluate these hypotheses and quantify the multisensory performance, we designed two simple models of word recognition. The first model is a minimal-integration model, in the sense that the integration of auditory and visual cues occurs within the lowest possible level of interaction between both inputs (i.e., probabilistic combination). The second model is an optimal-integration model in which individual spatio- and spectrotemporal audiovisual cues are combined across modalities to minimize the amount of information required for correct word recognition. We fitted a model of optimal multisensory integration to the performance of NHs with masked auditory input (Fig. 3*D*). We then compared the performance of the model with all subjects' performance in two other conditions (CI users at T0 and NH subjects with vocoder; Fig. 3 *A* and *C*). We found that the model fitted very well the performance of CI patients, indicating that at T0 they integrate visuoauditory inputs as efficiently as NHs when their auditory input is degraded by white noise. However, the bisensory performance of NHs with simulated implants was far below the model performance levels (Fig. 3*C*). Thus, in contrast with CI users, NH subjects did not integrate their visuoauditory input optimally when this auditory input is lacking fine spectrotemporal structure. Furthermore, CI users tested 1 year postimplantation showed a significant improvement of both auditory and visuoauditory performance while keeping a constant high speechreading recognition level. When applying the model to the unisensory performance of CI users at 1 year (Fig. 3*B*), we found that the evolution of multisensory performance with practice could be entirely explained by their increased auditory performance. This finding suggests that, whereas visual and auditory inputs are integrated optimally from the start, a reorganization of auditory cortices, supporting a better capacity for dealing with distorted auditory inputs, is the main cause for the quasiperfect multisensory performance reached by CI users after 1 year.

## Discussion

This study provides a long-term evaluation that shows the impressive benefits of cochlear implantation regarding the recovery of speech recognition because profoundly deaf patients can reach high rates of performance for hearing speech during the first 6 months postimplantation. The present data confirm that a profound hearing loss induces the acquisition of strong speechreading abilities (6, 19, 24, 25), but they represent the first evidence that this skill remains unaffected by the recovery of the auditory functions provided by the neuroprosthesis. CI patients preserve a striking speechreading ability acquired during the period of deafness while they have reached optimal auditory recognition. We interpret this apparently paradoxical strategy developed by CI users as a strategy to maintain through the mechanisms of bisensory integration a high level of speech recognition in a disturbed noisy auditory environment. Previous studies have reported that the performance of CI patients is highly susceptible to noise (16, 26), which is probably due to the

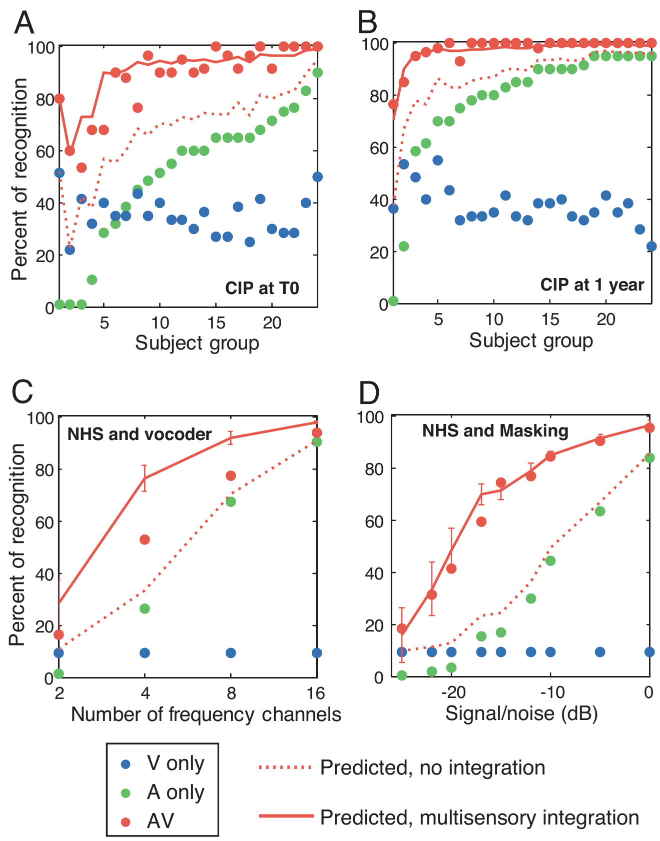



**Fig. 3.** Fitting a model of multisensory integration to the data. Filled circles represent the data (green, auditory; blue, visual; red, bisensory), dotted lines represent the predicted bisensory performance in the absence of multisensory integration ("model 1" in *Materials and Methods*), and plain lines represent the prediction of an optimal multisensory integration model with a threshold $T = 6$ ("model 2" in *Materials and Methods*). (*A* and *B*) CI users performance at T0 (*A*) and 1 year after implantation (*B*). Each group represents the averaged performance of three CI users ranked by their auditory performance. Thus, group 1 represents the (true and predicted) performance of the three subjects with the worst auditory performance, and group 26 represent the performance of the three subjects with the best auditory performance. (*C*) NH subjects' word recognition scores as a function of the amplitude of auditory white noise. (*D*) Word recognition scores of NH subjects as a function of the number of auditory frequency channels (simulated implants). This analysis shows that visuoauditory performance of CI users is perfectly in agreement with the optimal multisensory model,whereas NH subjects listening to CI simulations obtain bisensory performance far below that predicted by the integrating model.

lack of fine spectrotemporal information provided by the CI (16). In addition, as a consequence of multisensory integration (27), we show that speech intelligibility of CI users is greatly improved in visuoauditory conditions, especially during the first months postimplantation when auditory performance has not yet reached a quasioptimal level. The analysis of the performance of NH subjects submitted to degraded auditory stimuli and CI users confirmed that bisensory integration of visual and auditory speech information leads to improvement in speech intelligibility (28). However, the perceptual bisensory gain in NH subjects does not reach the same level as for CI users when compared at equivalent auditory performance. Although CI users are able to integrate their visuoauditory signal efficiently and compensate for the loss of spectral information, none of the naïve NH subjects listening to CI stimulations reach the same level of VA supraadditive integration. Altogether, we suggest that CI users have developed specific visuoauditory skills that lead to a powerful utilization of the visual spatiotemporal cues (29) provided by the lip and face movements (10), allowing these patients to reach near-perfect performance in visuoauditory situations. Using our computational model that allows us to avoid ceiling effects in subjects' performance, we confirmed that the performance of CI patients derived not only from higher efficiency in speechreading, but also from the acquisition of a higher skill level in multisensory integration when visual speech information is matched to an impoverished auditory signal.

Our results provide crucial information on the temporal window during which plastic changes can occur in the cortical network of CI patients during adaptation to the neuroprosthesis. There is now a growing body of evidence showing that sensory deprivation from early developmental stages has an important effect on the remaining sensory modalities (30, 31) through active cross-modal neuroplastic mechanisms (32, 33). In general, sensory deprivation leads to a compensatory increase in specific skills of the spared modalities that can be observed at both behavioral and neural levels in animal and human subjects (34, 35). However, cochlear implantation constitutes a unique approach to understand the cortical mechanisms that underlie the functional recuperation of the lost sensory modality. First, it has been shown that because CIs provide only a degraded signal that requires specific compensatory strategies, CI users present different levels of activation in auditory areas involved in semantic and/or phonological speech processing (36). Our longitudinal study on a large sample of patients suggests that such changes probably occur during the first 6 months depending on subjects' performance in speech recognition (37–39) and might remain different from normally hearing listeners (40) even after several years of auditory function recovery. Second, our results highlight that CI users develop a strong visuoauditory perceptive strategy for speech intelligibility while experiencing the reduced spectrotemporal information provided by the implant. This adaptation extends over the first 3 months postimplantation before being stabilized, suggesting that the pattern of brain activity during bisensory speech processing in CI users may vary during the corresponding period. Brain-imaging studies in CI deaf subjects have revealed a particular involvement of the low-level visual areas when listening to words (37). This finding corroborates our results of a strong synergy between visual and auditory processing for speech recognition following cochlear implantation. These results, in agreement with our ongoing functional imaging study (41), suggest that the visual activity derived from speechreading could actively influence the activity of the cortical network involved in hearing speech recognition and could participate in the improvement of performance in bisensory conditions. The existence of heteromodal connections that link directly unisensory areas in adult primates (42, 43) provides a possible anatomical framework for such direct visuoauditory interactions at low levels of sensory processing (44).

First, at a theoretical level, it has been shown that the fine spectrotemporal auditory information provides important cues for auditory speech recognition (45, 46). Our results broaden the role of the temporal fine structure because it optimizes the audiovisual speech integration leading to a higher multisensory perceptual benefit, in agreement with the actual technological challenge aiming at improving the spectral resolution of CIs. Second, from the clinical point of view, this work provides important cues to adapt the rehabilitation strategy as a function of implant experience. The supranormal skills in multisensory integration observed in CI deaf patients should be used to improve recovery of other auditory functions that are still deficient in CI users. Because visuoauditory training facilitates perceptual learning in a single modality (47, 48), we believe that a strong visually and audiovisually based rehabilitation during the first months postimplantation would significantly improve and fasten the functional recovery of speech intelligibility or sound localization, which is largely deficient in unilateral CI patients (49, 50).

## Materials and Methods

**Participants.** Our study was based on a retrospective analysis of speech recognition in 97 postlinguistically deafened subjects (mean age 56 years, range 19–82) that received a CI after profound deafness (defined as a hearing loss of ≥90 dB) of diverse etiologies (meningitis, chronic otitis, otosclerosis, neurinoma) and durations (mean age 22 years, range 1–57). The clinical implantation criteria included word and open-set sentence auditory-recognition scores <30% under best-aided conditions (i.e., with conventional acoustic hearing aids). All CI patients were recipients of a Nucleus (Cochlear) implant (CI-22 or CI-24) and used a range of different sound-coding strategies. Performance was collected during regular visits to the ear-nose-throat department following a standard rehabilitation program. We restricted our analysis to evaluations performed by the same speech therapist during the 10 years of follow-up and by using exactly the same procedures. First, we collected the performance of CI users tested before the cochlear implantation and by using an external hearing aid. On average, these tests were performed ≈6 months before the implantation (mean 5.8 months), but on 36 CI users of 97, word-recognition performance was obtained during the last 3 months preimplantation. Then from the day the CI was switched on (T0, usually 1 month postsurgery), CI users were tested at regular intervals during the first year and >8 years postimplantation. Data have been pooled into 12 groups (Fig. 1*A*) corresponding to the testing period from T0 ($n$ = 91), 3 months ($n$ = 91), 5 months ($n$ = 82), 7 months ($n$ = 77), 1 year ($n$ = 78), 2 years ($n$ = 69), 3 years ($n$ = 41), 4 years ($n$ = 26) 5 years ($n$ = 17), 6 years ($n$ = 11), 7 years ($n$ = 5), and >8 years ($n$ = 4) postimplantation. On average, CI users were tested during a period of 33 months postimplantation (± 25), with on average eight sessions per subject (±3). In this postlinguistic deaf adult population, we did not find a relationship between the age of implantation and the performance collected in the uni- or bisensory conditions (all cases $P > 0.05$). In addition, speech recognition for different sensory modalities was tested in a sample of 163 NH subjects. These control subjects were all native French speakers with self-reported normal or corrected-to-normal vision and without any previously known language or cognitive disorders.

**Procedures and Stimuli.** All subjects were tested on open-set recognition for French disyllabic words obtained from the classically used French speech therapist list developed by Fournier and presented in visual-only (speechreading, V), auditory-only (A), and visuoauditory (VA) conditions. Only words correctly repeated verbally by subjects were treated as correct responses (% correct score). We calculated the visual contribution to speech recognition by using the method of Sumby and Polack (8) (VA benefit = [(VA−A)/(100−A)]) to normalize for the performance observed in the A condition and thus to be able to compare directly the visuoauditory gain across groups (19). CI users were tested in silence on 20 words in each condition. In NH subjects, we developed three paradigms during which the auditory stimuli (the words pronounced by the speech therapist and recorded onto a PC computer) were differently degraded or presented without alteration. In a masking protocol, in A and VA conditions we additively combined each sound to a masking sound, with the words' acoustic level shifted to obtain the required SNR (nine SNR conditions: 0, −5, −10, −12, −15, −17, −20, −22, and −25 dB). The mask was a white noise delivered by a pseudorandom number generator and temporally modulated by monoperiodic sinusoidal lobes (period = 20 msec), with a mean rate of 300 modulations per sec temporally randomly distributed. Gain was 1 at the edge of the lobes and 0.4 at the center. This white noise modulation was carried out to ensure high random temporal fluctuations. Subjects ($n$ = 80) were tested with four lists of 20 words in A or VA conditions, with masking noise at a single SNR condition and the orders for each A or VA sequence being randomized across subjects. At each masking conditions in the range from −5 and −22 SNR, data were obtained from a sample of 10 subjects (but only 3 subjects at SNR 0 dB and 7 at SNR −25 dB). In a "simulating" protocol used with a second group of NH subjects ($n$ = 41), for A and VA conditions, we developed noise-band vocoder methods that simulate the signal processing computed in a CI (21). The sound was analyzed through 2, 4, 8, or 16 frequency bands by using sixth-order IIR elliptical analysis filters. The cutoff frequencies of these bands were calculated to ensure equidistance of the corresponding basilar membrane locations of the cochlea according to the human cochlear tonotopic map (51). Spectral analysis was systematically carried out between 250 and 8,000 Hz. Cutoff frequencies were 250, 1,676, and 8,000 Hz for the 2-band condition; 250, 709, 1,676, 3,713, and 8,000 Hz for the 4-band condition; 250, 437, 709, 1,104, 1,676, 2,507, 3,713, 5,462, and 8,000 Hz for the 8-band condition; and 250, 335, 437, 561, 709, 888, 1,104, 1,363, 1,676, 2,053, 2,507, 3,054, 3,713, 4,506, 5,462, 6,613, and 8,000 Hz for the 16-band condition. For each filtered frequency band signal, temporal envelope was extracted by half-wave rectification and envelope smoothing with a 500-Hz low-pass third-order IIR elliptical filter. The extracted temporal envelope was then used to modulate white noise delivered by a pseudorandom number generator, and the resulting signal was filtered through the same sixth-order IIR elliptical filter that was used for the frequency band selection. Finally, signals obtained from each frequency band were recombined additively, and the overall acoustic level was readjusted to match the original sound level. The performance of at least 10 subjects was analyzed for each band condition. In a last protocol, NH subjects ($n$ = 42) were tested on three lists of 20 disyllabic words presented in V conditions.

In all conditions, the lists of words were equalized for syllabic structure (CV/CVC/CCV), language utilization frequency (Brulex), and anterior–posterior phonemic constitution. The stimuli were uttered by the female French speech therapist, who pronounced each word with even intonation, tempo, and vocal intensity. Utterances were recorded in an anechoic chamber with a professional digital video camera with lights focused on the face such that minimal shadowing occurred. Video was digitized at 720 × 576 pixels at 25 frames per sec, and sound was digitized at 32,000 Hz by using a 16-bit quantization. Audiovisual stimuli with sound degradation were made by using Adobe Premiere Pro 7.0 (Adobe Systems, Mountain View, CA), and temporal coincidence was respected between the original and processed sounds. All stimuli were finally exported in MPEG2 video format with maximum encoding quality.

**Visual and Auditory Integration Models.** An increase in multisensory performance does not necessarily prove that subjects integrate their visuoauditory inputs. Indeed, being in the presence of two signals rather than one automatically increases the probability of recognition because a word can be recognized from one or the other signal. In model 1 (minimal integration), we suppose that the subjects recognize a word from what they see or hear, with two distinct and independent modules. In this case, the word will not be recognized if and only if both the auditory and visual modules fail to recognize it. If we call $P_{\mathrm{A}}$ the probability of recognizing a word from the auditory module and $P_{\mathrm{V}}$ the probability of recognizing a word from the visual module, then the probability of failure in the presence of both visual and auditory sensory input is $(1 - P_{\mathrm{A}})(1 - P_{\mathrm{V}})$. From this latter result, we can conclude that the probability of word recognition in the presence of the two modalities without integration $\hat{P}^{(1)}_{\mathrm{VA}}$ is given by $\hat{P}^{(1)}_{\mathrm{VA}} = P_{\mathrm{V}} + P_{\mathrm{A}} - P_{\mathrm{V}}P_{\mathrm{A}}$. Model 1 fails to account for the multisensory enhancement in performance in all conditions (red dotted line in Fig. 3). From this latter result, we can conclude that the visual and auditory modalities are indeed combined in a word-recognition task, albeit to a weaker extent in the case of NH subjects listening to the vocoder simulation.

To quantify multisensory integration for word recognition, we used a very simplified model where a word is "recognized" when a sufficiently large number of "cues" specific to this word is detected, either from the visual or auditory input (model 2: optimal multisensory integration). For example, a threshold of six means that six or more specific cues need to be detected to identify a word. These cues could be a specific motion of the mouth or a particular pattern in the time/frequency spectrum of the auditory signal (they do not necessarily correspond to phonemes). Moreover, we assume that the quality of the sensory input controls the average number of "cues" that can be detected in this condition.

The detection of each cue is probabilistic. We suppose that each cue is detected with a particular probability $P$ independently of the other cues. The resulting average number of detected cues is $\lambda = NP$, where $N$ is the total number of cues present in the word and $P$ is the probability of detecting each of them. $P$ depends on the quality of the sensory signal and controls the performance of the model. If $N$ is sufficiently large and $P$ is sufficiently small, the number of detected cues on each trial, $n$, follows approximately a Poisson law: The probability of detecting $k$ cues becomes $\Pi_\lambda(n = k) = (\lambda^k e^{-\lambda})/k!$. The probability of recognition corresponds to the probability that the number of detected cues will exceed a particular fixed threshold, i.e., $\Pi_\lambda(n > T)$. Thus, it is a function of both $\lambda$ and $T$. Assuming a fixed threshold $T = 6$ for each subject and condition, we can infer the mean number of cues $\lambda_V$, $\lambda_A$, and $\lambda_{VA}$ and detect the V, A, and VA sensory inputs for each subject in each condition. For example, $\lambda_A$ is the value for which $P_A = \Pi_{\lambda_A}(n > T)$ is equal to the observed auditory performance.

If visual and auditory inputs were combined optimally, they should add up together (the total number of detected cues is the sum of the visually and orally detected cues). Thus, the total signal should follow a Poisson law with mean $\hat{\lambda}_{VA} = \lambda_V + \lambda_A$. From this, we can infer the performance of an ideal observer: The bisensory recognition probability corresponds to the probability that a Poisson-distributed signal with mean $\hat{\lambda}_{VA}$ will exceed the threshold $T$. Thus, in the case of optimal multisensory integration, we have $\hat{P}^{(2)}_{VA} = \Pi_{\lambda_V+\lambda_A}(n > T)$.

Model 2 has one "free" parameter, the threshold $T$. The higher the $T$, the stronger the multisensory enhancement compares to unisensory performance. For $T = 1$, model 2 is equivalent to model 1 (there is no true integration: A word is detected by one or the other sensory modality). $T = 6$ was used for generating the model predictions because it is the best match for both the NH subjects in listening to noise-masked speech (Fig. 3*D*) and CI users (Fig. 3 *A* and *B*). By "best match," we mean that $T = 6$ minimized the mean squared error between the model predictions and the subject-per-subject performance in these conditions. Although the presence of a free parameter prevents us from proving that multisensory integration is optimal in an absolute sense, it provides a rigorous comparison between multisensory integration performance for different conditions and subject groups.

We thank M. L. Laborde and D. Sauvajon for help collecting the data; and C. James, S. Thorpe, and F. Polleux for comments and corrections on the manuscript. This work was supported by grants from the Action Concertée Incitative Neurosciences Intégratives et Computationnelles (to J.R., O.D., and P.B.), the Fondation pour la Recherche Médicale (to S.L.), and the Action Thématique et Incitative sur Programmes et Équipes program of the CNRS (to P.B.).

1. De Gelder B, Bertelson P (2003) *Trends Cogn Sci* 7:460–467.
2. Welch RB, Warren DH (1986) in *Handbook of Perception and Human Performance*, eds Boff KR, Kaufman L, Thomas JP (Wiley, New York), Vol 1, pp 1–36.
3. McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) *Nature* 407:906–908.
4. Vroomen J, de Gelder B (2000) *J Exp Psychol Hum Percept Perform* 26:1583–1590.
5. Stein BE, Meredith MA (1993) *The Merging of the Senses* (MIT Press, Cambridge, MA).
6. Summerfield Q (1992) *Philos Trans R Soc Lond B Biol Sci* 335:71–78.
7. McGurk H, MacDonald J (1976) *Nature* 264:746–748.
8. Sumby WH, Pollack I (1954) *J Acoust Soc Am* 26:212–215.
9. Helfer KS, Freyman RL (2005) *J Acoust Soc Am* 117:842–849.
10. Bernstein LE, Auer ET, Takayanagia S (2004) *Speech Communication* 44:5–18.
11. Schorr EA, Fox NA, van Wassenhove V, Knudsen EI (2005) *Proc Natl Acad Sci USA* 102:18748–187450.
12. Copeland BJ, Pillsbury HC, III (2004) *Annu Rev Med* 55:157–167.
13. United Kingdom Cochlear Implant Study Group (2004) *Ear Hear* 25:310–335.
14. Rauschecker JP, Shannon RV (2002) *Science* 295:1025–1029.
15. Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) *Nature* 352:236–238.
16. Fu QJ, Shannon RV, Wang X (1998) *J Acoust Soc Am* 104:3586–3596.
17. Brozinsky CJ, Bavelier D (2004) *Brain Res Cogn Brain Res* 21:1–10.
18. Neville HJ, Lawson D (1987) *Brain Res* 405:268–283.
19. Kaiser AR, Kirk KI, Lachs L, Pisoni DB (2003) *J Speech Lang Hear Res* 46:390–404.
20. Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2006) *Cereb Cortex*, in press.
21. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) *Science* 270:303–304.
22. Cienkowski KM, Carney AE (2002) *Ear Hear* 23:439–449.
23. Stanford TR, Quessy S, Stein BE (2005) *J Neurosci* 25:6499–6508.
24. Grant KW, Walden BE, Seitz PF (1998) *J Acoust Soc Am* 103:2677–2690.
25. Tyler RS, Parkinson AJ, Woodworth GG, Lowder MW, Gantz BJ (1997) *J Acoust Soc Am* 102:508–522.
26. Munson B, Nelson PB (2005) *J Acoust Soc Am* 118:2607–2617.
27. Bernstein LE, Auer ET, Moore JK (2004) in *The Handbook of Multisensory Processes*, eds Calvert G, Spence C, Stein BE (MIT Press, Cambridge, MA), pp 203–223.
28. Ross CA, Pearlson GD (1996) *Trends Neurosci* 19:171–176.
29. Schwartz JL, Berthommier F, Savariaux C (2004) *Cognition* 93:B69–B78.
30. Fieger A, Roder B, Teder-Salejarvi W, Hillyard SA, Neville HJ (2006) *J Cogn Neurosci* 18:149–157.
31. Gougoux F, Lepore F, Lassonde M, Voss P, Zatorre RJ, Belin P (2004) *Nature* 430:309.
32. Bavelier D, Neville HJ (2002) *Nat Rev Neurosci* 3:443–452.
33. Merabet LB, Rizzo JF, Amedi A, Somers DC, Pascual-Leone A (2005) *Nat Rev Neurosci* 6:71–77.
34. Roder B, Rosler F (2004) in *The Handbook of Multisensory Processes*, eds Calvert G, Spence C, Stein BE (MIT Press, Cambridge, MA), pp 719–747.
35. Rauschecker JP (1995) *Trends Neurosci* 18:36–43.
36. Giraud AL, Truy E, Frackowiak RS, Gregoire MC, Pujol JF, Collet L (2000) *Brain* 123:1391–1402.
37. Giraud AL, Price CJ, Graham JM, Truy E, Frackowiak RS (2001) *Neuron* 30:657–663.
38. Green KM, Julyan PJ, Hastings DL, Ramsden RT (2005) *Hear Res* 205:184–192.
39. Nishimura H, Doi K, Iwaki T, Hashikawa K, Oku N, Teratani T, Hasegawa T, Watanabe A, Nishimura T, Kubo T (2000) *Neuroreport* 11:811–815.
40. Demonet JF, Thierry G, Cardebat D (2005) *Physiol Rev* 85:49–95.
41. Laglеyre S, Rouger J, Laborde ML, Demonet JF, Fraysse B, Deguine O, Barone P (2006) in *Second Meeting of the European Societies of Neuropsychology* (Eur Soc Neuropsy, Boulogne Billancourt, France), p 162.
42. Cappe C, Barone P (2005) *Eur J Neurosci* 22:2886–2902.
43. Falchier A, Clavagnier S, Barone P, Kennedy H (2002) *J Neurosci* 22:5749–5759.
44. Schroeder CE, Foxe J (2005) *Curr Opin Neurobiol* 15:454–458.
45. Gilbert G, Lorenzi C (2006) *J Acoust Soc Am* 119:2438–2444.
46. Qin MK, Oxenham AJ (2003) *J Acoust Soc Am* 114:446–454.
47. Seitz AR, Kim R, Shams L (2006) *Curr Biol* 16:1422–1427.
48. Frassinetti F, Bolognini N, Bottari D, Bonora A, Ladavas E (2005) *J Cogn Neurosci* 17:1442–1452.
49. Grantham DW, Ashmead DH, Ricketts TA (2004) in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, eds Pressnitzer D, de Cheveigne A, McAdams S, Collet L (Springer, New York), pp 390–397.
50. Schoen F, Mueller J, Helms J, Nopp P (2005) *Otol Neurotol* 26:429–437.
51. Greenwood DD (1990) *J Acoust Soc Am* 87:2592–2605.