Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients

Pascal Barone, Laure Chambaudie, Kuzma Strelnikov, Bernard Fraysse, Mathieu Marx, Pascal Belin, Olivier Deguine

To cite this version:

Pascal Barone, Laure Chambaudie, Kuzma Strelnikov, Bernard Fraysse, Mathieu Marx, et al.. Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients. Cortex, Elsevier, 2016, 83, pp.259 - 270. 10.1016/j.cortex.2016.08.005 . hal-01469002

HAL Id: hal-01469002

https://hal-amu.archives-ouvertes.fr/hal-01469002

Submitted on 21 Feb 2017

Manuscript Number: CORTEX-D-16-00067R2

Title: Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients

Article Type: Research Report

Keywords: deafness; cochlear implantation; face; voice; audiovisual

Corresponding Author: Dr. Pascal Barone,

Corresponding Author's Institution: CNRS CERCO UMR 5549

First Author: Pascal Barone

Order of Authors: Pascal Barone; Laure Chambaudie; Kuzma Strelnikov; Bernard Fraysse; Mathieu Marx; Pascal Belin; Olivier Deguine

Abstract: Due to signal distortion, speech comprehension in cochlear-implanted (CI) patients relies strongly on visual information, a compensatory strategy supported by important cortical crossmodal reorganisations. Though crossmodal interactions are evident for speech processing, it is unclear whether a visual influence is observed in CI patients during non-linguistic visual-auditory processing, such as face-voice interactions, which are important in social communication. We analyse and compare visual-auditory interactions in CI patients and normal-hearing subjects (NHS) at equivalent auditory performance levels. Proficient CI patients and NHS performed a voice-gender categorisation in the visual-auditory modality from a morphing-generated voice continuum between male and female speakers, while ignoring the presentation of a male or female visual face. Our data show that during the face-voice interaction, CI deaf patients are strongly influenced by visual information when performing an auditory gender categorization task, in spite of maximum recovery of auditory speech. No such effect is observed in NHS, even in situations of CI simulation. Our hypothesis is that the functional crossmodal reorganisation that occurs in deafness could influence nonverbal processing, such as face-voice interaction; this is important for patient internal supramodal representation.

*Title page

Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients

Pascal Barone[1,2], Laure Chambaudie[1,2], Kuzma Strelnikov[1,2], Bernard Fraysse[3], Mathieu Marx[3], Pascal Belin[4,5], Olivier Deguine[1,2,3]

Shortened title: Face–voice interactions in cochlear-implanted patients

1. Université Toulouse, CerCo, Université Paul Sabatier

2. CNRS, UMR 5549, Toulouse, France

3. Service Oto-Rhino-Laryngologie et Oto-Neurologie, Hopital Purpan, Toulouse, France

4. Voice Neurocognition Laboratory, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

5. Institut de Neurosciences de la Timone, CNRS UMR 7289 et Aix-Marseille Université, Marseille, France

Corresponding author:

Pascal Barone

E-mail: pascal.barone@cerco.ups-tlse.fr

Centre de Recherche Cerveau & Cognition UMR 5549

Pavillon Baudot CHU Purpan

31062 Toulouse CEDEX9, France

Phone: +33 (0)5 62 17 37 79

Fax: +33 (0)5 62 17 28 09

***Highlights**

- We tested whether visual influence remains high in proficient CI deaf adults
- We presented a voice continuum with a male or female visual face
- CI deaf patients were strongly influenced by visual facial information
- No such effect is observed in controls, even for degraded auditory CI simulation
- Proficient CI deaf patients rely on visual cues during visuo-auditory perception

Abstract

Due to signal distortion, speech comprehension in cochlear-implanted (CI) patients relies strongly on visual information, a compensatory strategy supported by important cortical crossmodal reorganisations. Though crossmodal interactions are evident for speech processing, it is unclear whether a visual influence is observed in CI patients during non-linguistic visual-auditory processing, such as face–voice interactions, which are important in social communication. We analyse and compare visual-auditory interactions in CI patients and normal-hearing subjects (NHS) at equivalent auditory performance levels. Proficient CI patients and NHS performed a voice-gender categorisation in the visual-auditory modality from a morphing-generated voice continuum between male and female speakers, while ignoring the presentation of a male or female visual face. Our data show that during the face–voice interaction, CI deaf patients are strongly influenced by visual information when performing an auditory gender categorization task, in spite of maximum recovery of auditory speech. No such effect is observed in NHS, even in situations of CI simulation. Our hypothesis is that the functional crossmodal reorganisation that occurs in deafness could influence nonverbal processing, such as face–voice interaction; this is important for patient internal supramodal representation.

Key words: deafness; cochlear implantation; face; voice; audiovisual

1. Introduction

In profoundly deaf individuals, cochlear implant remains the most efficient solution to recover speech intelligibility and to restore social interaction to improve patient quality of life. However, sound processing performed by the implant provides only a crude signal that lacks fine spectral information. As a result, while CI affords acceptable levels of speech comprehension, spectral degradation impacts the ability of CI patients to process non-linguistic aspects of speech, such as changes in prosody and intonation (Green, Faulkner, Rosen, & Macherey, 2005; Marx et al., 2015; Peng, Chatterjee, & Lu, 2012) and most other voice features. Indeed, CI patients present deficits in discriminating human voice from environmental sounds (Massida et al., 2011), more specifically in recognising other voice attributes, such as gender, familiarity, or emotions of the speaker (Fu, Chinchilla, & Galvin, 2004; Fu, Chinchilla, Nogaki, & Galvin, 2005; Kovacic & Balaban, 2009; Massida et al., 2013).

However, in addition to the technical limitations of the implant processor, it should be considered that the brain reorganisation that occurs during deafness could be implicated in the global deficit present in cochlear implanted (CI) patients during voice processing as this has been proposed for auditory speech comprehension deficits (Lazard, Innes-Brown, & Barone, 2014). There is now compelling evidence that the success of CI for speech perception is highly dependent on the age at which the implantation is performed (Kral & O'Donoghue, 2010). This reflects the potential of brain plasticity, which is critical for the recovery of auditory function through the neuro-prosthesis during development (D. S. Lee et al., 2001; H. J. Lee et al., 2007), and even in adults (Strelnikov et al., 2015). Brain imaging studies point to a network of areas along the superior temporal sulcus and gyrus (STS and STG) that are specifically sensitive to human voice stimuli (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Kriegstein & Giraud, 2004; Pernet et al., 2015); This set of areas is referred to as temporal voice areas (TVAs) (Belin, Zatorre, & Ahad, 2002) and can be subdivided in various regions with distinct implications in human vocal sounds processing (Pernet et al., 2015) . In adult CI patients, TVAs are shown to be poorly activated by voice stimuli (Coez et al., 2008), a result that questions their functional integrity after a prolonged period of auditory deprivation. Further, it was demonstrated in deaf patients that the STS region, as part of the TVAs, is subject to crossmodal reorganisation during deafness. Firstly, it has been shown that the STS region responds to visual sign language in early deaf signers (Sadato et al., 2004) and similarly, the auditory TVAs are involved in visual speech processing through lip-reading in postlingual CI deaf patients (Rouger et al., 2012). In the cases of less severe hearing loss, there are also some indications of the take-over of the temporal auditory regions by visual functions (Campbell & Sharma, 2014). Lastly, there are numerous converging studies that demonstrate that the level of crossmodal reorganisation of the temporal auditory areas (STS and STG) is inversely related to the level of CI outcomes in young (H. J. Lee et al., 2007) and adult (Strelnikov et al., 2013) cochlear implanted deaf patients. While none of these studies provide evidence for a causal relationship between the cross-modal (visual) recruitment of temporal regions and deficient auditory processing in CI patients, these observations have been interpreted as a maladaptive impact of crossmodal reorganisation (discussed in (Heimler, Weisz, & Collignon, 2014)). Based on these interpretations, our hypothesis is that the TVAs have lost part of their functional integrity in CI patients (Coez et al., 2008), a phenomena that could be responsible to some extent for the deficit of CI patients in processing human voices and their attributes. While the crossmodal reorganization tends to decrease in the auditory temporal regions with the patients recovery of auditory speech comprehension (Chen, Sandmann, Thorne, Bleichner, & Debener, 2016; Doucet, Bergeron, Lassonde, Ferron, & Lepore, 2006; Rouger et al., 2012), we have no cues on how face-voice interactions are affected in CI patients. Recent evidence indicating that in CI users the auditory cortex responds to visual face stimuli (Stropahl et al., 2015) suggests that the integrative processing of the natural human face and voice stimuli could probably be different in CI deaf patients. Therefore, it is critical to assess how visual information can interfere with auditory voice processing in deaf CI patients.

The voice signal, considered as the auditory face (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin, Fecteau, & Bedard, 2004), carries speech information as well as non-speech identity information about gender, age, physical factors, and emotions. In addition, visual and vocal information about the speaker's state of mind shows strong complementarity, as paralinguistic (Foxton, Brown, Chambers, & Griffiths, 2004; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004) or affective information is also supported by crossmodal face–voice interaction (Collignon et al., 2008; de Gelder & Vroomen, 2000). Based on such strong complementarity, models of face–voice interactions have been proposed involving an internal supramodal representation of the person (Campanella & Belin, 2007). While visual-auditory interactions for speech comprehension have been substantially addressed in CI deaf patients (Barone & Deguine, 2011 for review), to our knowledge, the literature shows no indication on how facial information can influence voice processing. However, the recent observation of a crossmodal activation of the auditory cortex by visual face presentation in CI patients (Stropahl et al., 2015) suggests a probable impact of deafness on face-voice interactions. Concerning speech processing, CI patients rely strongly on visual-auditory interactions, because the visual information obtained from lip-reading allows disambiguation of the impoverished signal delivered by the implant and acts as an enhancement of the signal-to-noise ratio for speech comprehension in a noisy environment (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954). By analogy with what is reported for speech, our hypothesis is that bimodal face–voice interactions should be prominent in CI patients, in light of their difficulties in perceiving voice features, even after a long period of experience with the implant (Massida et al., 2011; Massida et al., 2013).

Furthermore, during speech processing, when there is a mismatch with the auditory signal, such as in the McGurk protocol (McGurk & Macdonald, 1976), CI patients are highly sensitive to visual information, and they tend to respond toward the visual modality (Desai, Stickney, & Zeng, 2008; Rouger, Fraysse, Deguine, & Barone, 2008), while normal-hearing subjects (NHS) fuse both types of information. However, it is unclear how the visual bias observed in CI patients is dependent on the level of auditory speech recovery (Rouger et al., 2008; Tremblay, Champoux, Lepore, & Theoret, 2010) as some previous studies showed that it is restricted to patients with low and medium auditory recovery (Champoux, Lepore, Gagne, & Theoret, 2009). As a result, in order to infer that abnormal face–voice interactions in CI patients are independent from the level of speech recovery with the CI , it is critical to analyse and compare visual-auditory interactions in CI patients and NHS at comparable performance levels in the auditory modality. To achieve this constraint, first we recruited highly experienced CI patients with at least one year of implant experience and with the criteria of presenting high performance level in speech comprehension. In addition, the impact of visual information on the auditory processing of such highly infrequent patients is further compared to NHS stimulated with distorted auditory information that simulates the processing of a CI.

We asked CI deaf patients to perform a voice-gender categorisation from a morphing-generated voice continuum between a male and a female speaker, while ignoring the presentation of a male or female visual face. We expected a strong visual influence from the visual modality in CI patients, an interaction that should be more robust than that observed for speech-based processing, on the assumption that face and voice information are automatically merged (Amedi, Malach, & Pascual-Leone, 2005). Furthermore, we asked if such strong face–voice interactions can be observed in experienced CI patients with a high level of recovery in speech-processing. Indeed, in the present study when compared to NHS undergoing a CI simulation, our results clearly demonstrate that experienced CI patients are much more sensitive to face information during incongruent visual-auditory situations. Such results represent further evidence that the predominant influence of the visual modality, in cases of conflicting or ambiguous multimodal conditions, is independent of the level of auditory speech recovery. We hypothesise that because CI patients rely strongly on visual-auditory synergy to process auditory information, they are more susceptible to visual interference. Lastly we proposed that this phenomenon could probably be supported by the functional crossmodal reorganisation of the auditory temporal areas, including the TVAs that occurs during deafness, a hypothesis that needs further investigation based on objective brain imaging data.

2. Materials and Methods

2.1 Participants

Normally hearing. A group of 32 native-French-speaking NHS (16 men, age 25 ± 7 mean ± SD) with no self-reported history of auditory, neurological, or psychiatric disorders participated in the study and performed a voice-gender categorisation task. Twenty-two of the 32 NHS were asked to perform the voice-gender categorisation task using only the original voices. The other 10 participants performed the task with a vocoding condition in addition to the original voices.

CI deaf patients. Fourteen CI deaf patients (age 61.71 ± 14 years mean ± SD; men) participated in the study. The cohort of CI patients is older than the set of control subjects, the latter having been selected to present optimal performances in processing auditory information. However, while aging can be an important issue on perceptive and cognitive functions (see (Baltes & Lindenberger, 1997)), it is important to mention that a large proportion (6/14) of the CI patients presented an age range below 60 years which cannot be considered as within the critical period for sensory decline. Nevertheless, we cannot exclude that part of the differences with NHS could be influenced by the age of the CI patients. In addition, as CI outcomes for speech depend also on various psycho-cognitive functions (see Francis et al 2014), we believe that these experienced CI patients have a high probability of presenting preserved cognitive functions in spite of their relative higher age.

All patients had postlingually acquired profound bilateral deafness (defined as hearing loss of ≥ 90 dB) of diverse aetiologies (see Table 1) and durations (11.51 ± 9 years mean ± SD). Clinical implantation criteria included word and open-set sentence auditory-recognition scores below 30% under best-aided conditions with conventional acoustic hearing aids. All CI patients received a Nucleus (Cochlear) implant (CI-22 or CI-24), with a range of different sound-coding strategies (ACE, SPEAK). These patients were carefully selected according to a criterion of successful auditory recovery in terms of speech comprehension post-implantation. We have arbitrarily chosen a selection criteria of above 65% of auditory speech comprehension based on our previous analysis regarding the dynamics of recovery (Rouger et al., 2007). These “expert patients” had been implanted for at least 1 year at the time of testing (CI experience 7.61 ± 9 years mean ± SD) and they showed optimal recovery of speech intelligibility with word or open-set sentence auditory recognition (86.07 ± 11% mean ± SD correct for auditory disyllabic words in quiet). The main criterion was the high proficiency of the patients in clinical speech-perception scores and the duration of at least one year post-implantation was chosen to ensure the stability of their good performance. For the present study, performance in voice-gender discrimination was measured during regular visits to the ENT department following a standard rehabilitation program. All patients gave their full, informed consent prior to participation in this study in accordance with the Declaration of Helsinki (1975), and institutional ethics committee approval was obtained (CPP Sud-Ouest et Outre Mer 1, n°08 261 03).

2.2 Stimulus and procedure

All stimuli were developed at the Voice Neurocognition Laboratory of the University of Glasgow (http://vnl.psy.gla.ac.uk). We used a subtest of the Voice Perception Assessment (VPA) battery (Pernet and Belin 2012, see http://experiments.psy.gla.ac.uk/experiments/assessment.php?id=35) that was used in a study on voice-gender performance in CI patients (Massida *et al.*, 2013). The task requires participants to categorise by gender voice stimuli from a morphing-generated voice continuum between a male and a female voice speaking the same syllable “had”. The two-extreme voices each correspond to an average voice from 16 voices of the same gender. Morphing was performed using STRAIGHT (Hideki Kawahara, University of Wakayama) (31) in Matlab 6.5. STRAIGHT performs instantaneous pitch-adaptive spectral smoothing to separate the contributions of the glottal source (including F0) from the supra-laryngeal filtering (distribution of spectral peaks, including the first formant F1) to the voice signal. Voice stimuli are decomposed by STRAIGHT into five parameters: fundamental frequency (F0), formant frequencies, duration, spectro-temporal density, and aperiodicity; each parameter can be independently manipulated. Anchor points, that is, time-frequency landmarks, were determined in both extreme voices based on easily recognisable features of the spectrograms. The temporal landmarks were defined as the onset, the offset, and the initial burst of the sound. Spectro-temporal anchors were the first and second formant at onset of phonation, onset of formant transition, and end of phonation. Using the temporal anchors, elements of the continuum were equalised in duration (392 ms long, *i.e.*, 17,289 data points at

44.1 Hz). Morphed stimuli were then generated by re-synthesis based on a logarithmic interpolation of female and male anchor templates and spectrograms in steps of 10%. We thus obtained a continuum of 11 voices ranging from 100% female to 100% male with 9 gender-interpolated voices in 10% steps.

Based on these stimuli, we created a vocoded condition of only two channels (see Rouger *et al.*, 2007; Massida *et al.*, 2011 for the vocoding procedures). The sound was analysed through two frequency bands by using sixth-order IIR elliptical analysis filters. For each filtered frequency band signal, the 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 used for 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.

Two conditions of voice-gender categorisation were conducted. In a first condition, the test was presented in the auditory modality alone (A). In this case, the participants were asked to categorise the voices as male or female. In a second audiovisual (AV) condition, the auditory stimuli were paired to a male or a female static face presented on a monitor. The auditory stimuli were centred on the 1,500 ms period of face presentation, leaving a short period of 550 ms of face presentation alone. This sequence of presentation was chosen in order to be comparable with a previous study that used non-linguistic and linguistics visual stimuli (color, moving dots, lip motion) to analyze visuo-auditory interactions in CI users (Champoux et al., 2009). From this face/voice pairing, we obtained five congruent AV simulations (AVc) in which a male (or female) voice was presented with a male (or female) face. Conversely, we obtained five incongruent face–voice associations (AVic). Visual stimuli consisted of two colour photographs of a male and a female that we selected to be highly representative, as these faces were 100% categorised as male or female by 10 extra subjects. The stimuli were reworked using Adobe Photoshop and were normalised for light and contrast using Matlab 6.5.

Subjects were tested in a sound-attenuated chamber with volume adjusted to 72 dB SPL. NHS were tested at the CerCo Laboratory and CI patients at the Purpan, Toulouse. Auditory stimuli (16-bits, stereo, 22,050 Hz sampling rate) were presented binaurally via Sennheiser Eh 250 headphones.

First, CI patients and NHS were tested in the AV condition, during which the 220 face–voice–paired stimuli were randomly presented (22 face–voice combinations repeated 10 times). After a rest, the participants were asked to categorise voice-gender in an A condition (11 voices repeated 10 times) presented in random order. A set of 10 out of the 32 NHS performed the A and AV tasks with the two-channel vocoding condition added.

Participants were asked to perform a forced-choice gender categorisation, focusing their attention on the auditory input rather than the face. NHS were tested with a 1 s inter-trial delay (between the response and the new presentation), with the instruction to respond as quickly and accurately as possible using the left or right control buttons of the computer keyboard corresponding to their answer (male or female). CI patients were tested with a 1.5 s inter-trial delay and were instructed to answer as accurately as possible, with no reference to reaction time. The response keys were counterbalanced across subjects.

2.3 Data analysis

We calculated the rate of responses “female” for each of the 11 voices, from the male voice to the female voice. A Boltzmann sigmoidal function was fitted to the response points using a non-linear least-squares procedure (Levenburg-Marquart algorithm, Origin v6.1) to evaluate the categorisation response on multiple criteria (see Massida *et al.*, 2013, Fig 1A). First, we measured, on a curve, the stimulus for which the subjects present a chance level response (50%) corresponding to the ambiguous voices (C50 threshold). Second, we measured the slope at this point of the curve. Third, we analysed the percentage of correct gender recognition at the extremes for the unambiguous voice stimuli.

To analyse the effect of simultaneous presentation of a visual face on voice categorisation performance, we computed, for each subject, a Visual Impact index (VIx), obtained from the psychometric functions in the A and AV conditions (Fig 1A). First, we calculated the area under the curve (AUC) of the psychometric functions separately for each side (male or female) of the continuum. This computation was made in A, AVc, or AVic conditions. The values were standardised through a mirror image with respect to the response rating code (male or female) to be comparable between each side of the continuum. VIx corresponds to the ratio of the surface area obtained in A and AV conditions normalised with respect to the A conditions (VIx = AV-A/A), the AV conditions (female and male face) are averaged. If the presentation of a visual face influences the voice gender categorisation, we should observe an increase of the AUC representing a facilitator effect in case of congruency. This would increase the VIx values. Inversely, in the case of incongruence between the voice and face stimuli, the AUC would be reduced as well as the visual index. Values close to zero indicate an absence of influence of face presentation on auditory categorisation. In addition, a shift of the sigmoid function toward the visual stimuli should also be expressed as a shift in the C50 Threshold.

Direct comparisons of the performances (VIx, slope values) between groups were performed using the bootstrap method with bias-corrected and accelerated confidence intervals (Carpenter & Bithell, 2000) because these values were not normally distributed. For the same reason, non-parametric Spearman rank correlations with VIx were used in the analysis. The data for each group were re-sampled 10,000 times to obtain a distribution of 10,000 stimulated observations and the mean of the sample. On the basis of this simulated distribution, the effect was considered significant if there was no overlapping of confidence intervals at $p < 0.05$. Since comparison of values that do not differ from zero would not be informative, we first tested for a significant deviation from zero amongst the VIx and slope values (corrected alpha level = 0.0083). In the case of absence of significant difference, we provide the uncorrected p-values 0.05 for bootstrap as a more liberal threshold. All data are presented in the results using the mean and the confidence interval of the mean (in brackets).

3. Results

Because CI recipients rely strongly on visual cues, our goal was to assess the influence of the presence of a face on auditory voice categorisation, with the expectation of a bias in CI patients toward information provided through the visual modality. First, we compared the performances of the expert CI patients in A and AV conditions to compute a VIx, before comparing these VIx values to those obtained in NHS stimulated with an original or vocoded voice stimulus.

3.1 Voice-gender categorisation in original conditions

Figure 1B illustrates the psychometric function of the NHS during the A-only categorisation task. As shown in a previous study (Massida et al., 2013), subjects categorised correctly the unambiguous voice at the extremities. When stimuli were closer to the androgynous voice (50% on the continuum), subjects categorised the voice as female half of the time and male the other half. Globally, the psychometric curves of the participants can be fitted with a sigmoid function.

The CI patients were selected based on a very high level of auditory speech comprehension recovery following at least 1 year of experience with the implant. However, while we did not have any prediction on their performances, they presented an impressive outcome close to normal performance in categorisation (Figure 1d). None of the parameters derived from the psychometric functions were significantly different between NHS and CI patients at corrected and uncorrected levels (slope: CI: 0.34 [0.26;0.41], NHS: 0.64 [0.36;1.69], n.s., $p < 0.05$; AUC – “Male” side: CI: 0.67 [0.5;0.9], NHS: 0.65 [0.51; 0.83]; AUC “Female” side: CI: 4.07 [3.52; 4.32]; NHS 4.31 [4.09; 4.46], n.s., *p uncorr.* < 0.05). Therefore, it was possible to compare visuo-auditory interactions in CI patients and NHS in subject groups presenting an equivalent performance level of voice-gender categorisation in the auditory modality alone. However, it is important to keep in mind that the results obtained in the “expert patients” are not representative of the general weaker performance in voice-gender discrimination observed in larger populations of CI users (Kovacic & Balaban, 2009; Massida et al., 2013).

For the NHS in the AV conditions, during which the task was to concentrate on the auditory stimuli only, simultaneous presentation of the visual face had no effect on performance. As shown in Figure 1b, the performances of the NHS were very similar in A and AV conditions, irrespective of the gender of the face presented and of the congruency condition (AVc and AVic). Consequently, the visual index (VIx) values were not significantly different from zero at the corrected and uncorrected levels in the AVc (0.01 [-0.01; 0.03], *p uncorr.* < 0.05) and AVic (-0.014 [-0.04; 0.01], *p uncorr.* < 0.05) conditions.

In contrast, in CI patients, we observed a strong influence from presentation of the face on auditory categorisation. As shown in Figure 1d, when a face was simultaneously presented, the psychometric function was notably shifted toward the gender carried by the face, particularly when the face was incongruent with the voice (*e.g.*, a male face paired with a voice on the female side of the continuum). This effect is expressed as a decrease in slope values in AV compared to the A condition (0.15 and 0.17 in AV vs. 0.22 in A-only). Furthermore, we noted a decrease in gender recognition performance for the extreme unambiguous voices in AVic conditions compared to the auditory condition (15% [6; 36] vs. 3.7% [0; 6]), respectively, *p uncorr.* < 0.05).

Thus, visual impact (Figure 2) that takes into account performance across the continuum is significantly different from zero only in patients for the AVic presentations (-0.19 [-0.06; -0.50], p< p=0.0083). However, CI patients present a strong inter-individual variability in the VIx values (see Sup. Figure 1) and 3 of them present a large influence of the incongruent visual face. Still a significantly negative value of VIx persisted even after excluding, from the analysis, the three patients with the most negative VIx (-0.06 [-0.02; -0.16], p< p=0.0083). Such result suggests that, in spite of a large variability, the results are still resistant to the exclusion of possible outliers. It is important to mention that the 3 patients with the most extreme VIx values were the youngest patients of the CI group, an observation that is in favour of the absence of an effect of aging for the higher visual bias observed in CI users. However, since the two groups differed not only in terms of hearing status but also in terms of age, a potential confounding effect of age cannot be excluded. In the congruent AVc conditions, VIx was not different from zero even at the uncorrected level (0.05 [-0.03;0.25], *p uncorr.* < 0.05). Again, the values were also quite heterogeneous in this congruent AV condition and one CI patient presented a large VIx suggesting a strong facilitating influence of the visual facial stimulus during a congruent AV presentation. However, at the population level such effect was not significant.

First, these data reveal a dichotomy in the impact of presentation of facial information on auditory voice processing. The significantly negative values of VIx signify that a deleterious effect is present in the AVic situation. However, that VIx does not differ from zero in the AVc condition (bootstrap) suggests no facilitatory influence of the visual information on voice-gender categorisation when the auditory and visual stimuli are semantically congruent.

Importantly, the bootstrap analysis confirms that the VIx is much more negative in CI recipients compared to NHS in the AVic (CI: -0.19 [-0.10;-0.33]) NHS: -0.014 [-0.04; 0.01], $p < 0.05$) condition but in the AVc condition both do not differ from zero. Thus, in CI patients, bimodal stimulation is deleterious in AVic condition. None of the effects (facilitatory or deleterious) are observed in NHS.

Because the task of gender categorization was a difficult task for CI patients (see (Massida et al., 2013)), they were tested with no specific instruction of speed and with constant ITDs in order to reduce the difficulty of the test. Further, we applied the same protocol as the one performed on a different and larger set of CI patients with variable CI duration exposure (Massida et al., 2013) to insure and validate the reproducibility of the results in the A-only condition. In this situation, there was no justification to compare the reaction times of patients and control subjects as CI patients were much slower to respond (1.16 s vs. 0.73 s respectively). However, it is interesting to mention that both groups present an increase of the RTs values when the voices are approaching the most ambiguous androgynous voice (see Sup. Figure 2) suggesting that CI patients have developed a similar behavioral response type. Interestingly, in CI users we observed a shortening of RTs in AV compared to A-only conditions (1.20 s vs. 1.12 s) but no such multisensory effect was present in NHS (see Sup. Figure 2). These results reinforce the hypothesis towards a higher sensitivity in CI patients to facial information while processing auditory voice.

3.2 Voice-gender categorisation by NHS in degraded condition

Our results clearly demonstrate a strong influence of face presentation on auditory voice categorisation in CI deaf patients. However, when considering the strong association between face and voice during personal identity processing, one might plausibly interpret such effect of face in CI deaf patients as resulting from an imbalance in favour of the visual channel because of the degraded information delivered by the implant. Indeed, there is now a vast literature showing that multisensory interactions are prominent in situations that are approaching the perceptual threshold (Ross et al., 2007). To rule out this possibility and to attribute this effect, in CI users, to a functional adaptation induced by deafness, we compared the performances of the CI deaf patients to those obtained in NHS, tested through simulation of a CI processor (vocoding). This comparison has been efficient to demonstrate that CI users can present stronger audio-visual integration for speech (Rouger et al., 2007) while it is limited by the fact that control subjects are naïve to the CI simulation.

In conditions of degraded auditory stimulation, NHS presented strong impairment in voice-gender categorisation, as illustrated in Figure 1c. The psychometric function presented a weaker slope compared to that obtained in the original voice condition (0.15 [0.11;0.19] vs. 0.64 [0.36;1.69], $p < 0.05$, bootstrap), and performance levels were much lower for the extreme unambiguous voices (88.3% [81.1;93.3] vs. 96.3% [93.7;100], $p < 0.05$, bootstrap). Such performances are close to those observed in inexperienced CI users in the first months after implantation (Massida et al., 2013).

Critically, when presented simultaneously with a visual face, NHS stimulated with a vocoded sound maintained a similar level of performance to that in the A condition. We did not find an effect of face presentation on the visual index values (Figure 2); at the group level, the visual Index values are statistically not different from zero even at the uncorrected level, a result that confirms the absence of any kind of effect of visual information on voice categorisation in both conditions (AVc 0.002 [-0.07;0.06], AVic 0.004 [-0.06;0.1] , *p uncorr.* <0.05).

In summary, in the AVic condition, CI patients present a significant visual impact (negative VIx) but no such impact exist in the NHS stimulated with a two-channel vocoder.

3.3 Correlation analysis

As explained above, our strategy was to select experienced CI patients who could perform optimally the voice categorisation task as suggested by their speech perception scores. Based on previous reports claiming that crossmodal interaction is dependent on the level of CI recovery or CI experience (Tremblay et al., 2010), we searched for any correlation between VIx values and patient history. Firstly, as the patients are older than the NH controls, we searched for an effect of age on the VIx values based on some assumption of specific effect of age on multisensory processing (Laurienti, Burdette, Maldjian, & Wallace, 2006; Mahoney, Li, Oh-Park, Verghese, & Holtzer, 2011). A recent study in elderly and young CI patients showed that multisensory integration is present at all ages and that, older CI patients tend to be more reactive to auditory stimuli than younger CI patients (Schierholz et al., 2015). Based on such observations, we performed a correlation analysis with age but we did not observe any correlation among individual VIx observed in AVic or AVc conditions with patient age (rho=0.51, p>0.06; rho=0.07, p>0.8). Similarly, negative results were obtained with the duration of deafness (rho=-0.07, p>0.8; rho=0.20, p>0.4), duration of CI experience (rho=0.13, p>0.6; rho=-0.30, p>0.27),), or performance on disyllabic word comprehension (rho=0.53, p>0.06; rho=-0.31, p>0.26) (Spearman correlation values first for VIx for AVic, then for AVc). Given the non-normal distribution of the data, we used Spearman correlation, which could result in a certain loss of sensitivity of the analysis, in particular due to outliers. We checked for outliers in these correlations using a criterion of 3 standard deviations and no particular points were detected according to this criterion. Thus, we did not find a significant dependency on patient history in terms of duration of auditory deprivation or the experience of the implant.

4. Discussion

Our data show that during face–voice interaction, CI deaf patients are strongly influenced by visual information when performing an auditory gender categorisation task, despite maximum recovery of auditory speech comprehension allowed by the neuro-prosthesis. No such effect is observed in NHS, even in situations of strong auditory degradation that mimic the low resolution of a CI processor. The study provides evidence of a visual bias in CI patients while they were asked to categorise gender identity based on conflicting and ambiguous audiovisual information even when asked to ignore visual information. However, there were no differences from controls in the auditory gender categorisation and no facilitation effects in the audio-visual congruent condition, which may be due to a certain ceiling effect in the auditory voice gender categorization. Our data demonstrate that visual interference with auditory processing affects the nonverbal domain and concerns the information contributing to personal recognition embedded in facial and vocal perception.

Because speech is by nature multisensory (Vatakis, Ghazanfar, & Spence, 2008), it is clearly demonstrated that CI deaf patients present atypical audiovisual interactions when this is related to language processing. CI patients present supra-normal skills of multisensory integration of speech (Rouger et al., 2007) and high proficiency in audiovisual fusion, due to persistent use of visual information derived from lip-reading to compensate the impoverished signal transmitted by the processor. As a result of the strong dependency on visual cues for speech comprehension, several studies show that when bimodal speech information is ambiguous, such as in the McGurck condition, CI patients base their perceptual decisions on their most reliable sensory channel, vision (Bayard, Colin, & Leybaert, 2014; Desai et al., 2008; Rouger et al., 2008). We demonstrate evidence that such visual bias occurs similarly for non-linguistic face–voice interaction. When CI recipients are engaged in a voice-gender categorisation, visual face stimulus influences gender perception specifically when the auditory information is ambiguous.

Human social relations rely strongly on face–voice interaction, and perception of a personal identity benefits from multimodal integration of facial and vocal information (see (Campanella & Belin, 2007). In normal individuals, perception of most of the information carried by a voice (emotion, gender or identity) can be modulated by simultaneous presentation of a face (Collignon et al., 2008; de Gelder & Vroomen, 2000; Latinus, VanRullen, & Taylor, 2010; Schweinberger, Kloth, & Robertson, 2011). Such bimodal interactions are expressed as a facilitation of voice perception when information from visual and auditory modalities is semantically congruent (Belin, Campanella, Ethofer, & Schweinberger, 2012). In the present protocol, during an AVic presentation, when control subjects are asked to ignore the visual information, gender categorisation is not influenced by visual stimuli, even with degraded vocoded auditory condition. These results could appear to be in contradiction with previous studies that have reported an influence of a face stimulus when attention is directed toward the voice (see (Latinus et al., 2010)). It is important to mention that the strength of face-voice interactions depends on the information to be categorized (gender, emotion, age…) and that a still image as presently used is probably not as vivid as a dynamic face to study face-voice interactions (Watson et al., 2013).

In addition, as a result of recruiting experienced CI patients with strong recovery for speech comprehension, we did not observe a significant deficit in the performance of patients in the auditory condition as we have previously observed in a large set of patient (Massida et al., 2013). However, CI patients can develop adaptive strategies to categorize natural sounds including human voices (Collett et al., 2016), an ability that tends to be improved with cochlear implantation experience. The comparable auditory performance between NHS and CI patients in the unimodal situation rules out the possibility that the visually biased decision present in CI patients could be due to the "inverse effectiveness" principle that characterises multisensory integration processing (see (Stein & Rowland, 2011)) for a review). This principle states that as the performance in a single sensory stimuli decreases, the strength of multisensory integration should increase. No such effect is observed in NHS during the CI simulation, a result that suggests that the visual modulation observed in CI patients is specific to deafness and recovery through the implant.

Our results agree with those reported in CI patients in a McGurk protocol (Desai et al., 2008; Rouger et al., 2008). Indeed, when bimodal speech information is ambiguous, such as for incongruent audiovisual places of articulation, CI patients tend to overweight visual cues compared to auditory cues (Rouger et al., 2008). In this case, CI patients rely more strongly on the visual channel, which they consider more reliable. The inclination of CI patients to be more confident in the visual channel is also evident when analysing a restricted set of data in which patients must determine the gender of the person with no specification on the sensory modality to be used in the task (see Sup Figure 3). The results are quite variable due to the small number of patients tested and by the fact that some subjects tend to respond mainly with respect to the voice while others with respect to the face. The small number of patients precludes making robust statistical analysis on this distribution. However on average, when face-voice are incongruent, even when the voice information is unambiguous (at the extremes), the patients tend to respond predominantly towards the gender carried by the face (the responses are more than 60% towards the gender that corresponds to the face). While NHS responses in this case are also biased towards the face, the responses tend to remain in majority towards the gender that corresponds to the voice. Such results can be interpreted as supplementary evidence of a higher sensitivity of CI patients to the visual information during the incongruent face-voice presentation. The auditory voice information remains the primary source of decision to categorize gender in NHS (see Watson and Belin unpublished) but not in CI patients, specifically in case of ambiguity.

Additionally, it has been proposed that crossmodal interference is dependent on the level of recovery, as there is an inverted impact of visual interference in auditory speech processing with respect to the level of CI proficiency (see (Voss, Collignon, Lassonde, & Lepore, 2010). We did not test neo-implanted patients, because their performance in categorising the voice-gender is low (Massida et al., 2013). Nevertheless, we expect that in non-proficient CI patients who present a strong deficit in voice-gender categorisation, the visual influence would be much stronger. However, our results contradict previous studies (Champoux et al., 2009; Tremblay et al., 2010) claiming that proficient CI patients present normal integration of incongruent visuo-auditory information. Here, we show that in spite of a strong speech comprehension recovery, coupled with a near-normal ability to perform the auditory categorisation, our selected cohort of "expert" CI patients is strongly influenced by the face stimulus. This is further reinforced by the lack of correlation between the strength of visual integration and the duration of CI experience or the level of speech comprehension recovery. Such apparent discrepancy could be due to the type of stimuli used to assess visuo-auditory interactions as in the previous report, linguistic and non-linguistic visual cues can impact multimodal interactions differently in CI users (see (Champoux et al., 2009)). Unlike most complex multimodal objects, face and voice information are reflexively merged together (Amedi et al., 2005), a particularity that may explain the reminiscent susceptibility of voice processing to facial information in experienced CI patients. Indeed, a recent EEG study revealed that CI patients present a specific response to faces in the auditory cortex (Stropahl et al., 2015). Further, the amplitude of this crossmodal response in the auditory cortex is correlated to the performances of CI users in a face memory test.

The later results have been interpreted by Stropahl et al. (2015) as an adaptive cross-modal reorganization for processing visual information. But as stated in their article, there were no indications on how auditory processing was related to the visual takeover. However, we propose that these results reinforce our previous hypothesis that the visuo-auditory interactions observed in CI patients originate, at least partly, in the mechanisms of crossmodal reorganisation that occur during deafness and, progressively after CI. Face and voice processing share an analogous mechanism, while they are supported by separate neuronal structures (see (Yovel & Belin, 2013). The human brain presents specific cortical areas, in the occipito-temporal and superior temporal regions, that are more sensitive to human face or voice stimuli, respectively (Belin et al., 2000; Haxby, Hoffman, & Gobbini, 2000). Interestingly, both the fusiform face and temporal voice-selective areas (FFA and TVA, respectively) present crossmodal reorganisation following early blindness (FFA: (Gougoux et al., 2009; Holig, Focker, Best, Roder, & Buchel, 2014a, 2014b) or deafness (TVA, (Sadato et al., 2004) and they show an increase of functional coupling during explicit face–voice association learning (von Kriegstein & Giraud, 2006). The auditory areas of the STS/STG region globally showed different levels of crossmodal reorganisation, in congenital deaf patients (Sadato et al., 2005; Sadato et al., 2004; Vachon et al., 2013), in patients with hearing loss (Campbell & Sharma, 2014) and in CI patients (Doucet et al., 2006; H. J. Lee et al., 2007; Rouger et al., 2012; Song et al., 2015). In adult deaf CI patients, we demonstrate that auditory TVA is likewise the locus of crossmodal reorganisation, showing specific activation by visual speech information (Rouger et al., 2008; Rouger et al., 2012). Knowing that in NHS, the face- and voice-selective areas are functionally (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005) and structurally (Blank, Anwander, & von Kriegstein, 2011; Ethofer et al., 2013) directly connected, our hypothesis is that in CI patients such privileged intermodal connectivity (see (Joassin et al., 2011; Watson et al., 2014) is reinforced in spite of the auditory recovery, leading to a clear impact of visual face information on auditory gender categorisation. We expect that such reinforcement leads to an unconscious overweighting of visual face information in a face–voice binding that might be automatically processed through the connectivity of unimodal face- and voice-selective areas (Amedi et al., 2005).

Numerous factors are involved in the extent of crossmodal reorganisation during sensory loss including deafness (see (Voss et al., 2010) for review) among which the age, duration and severity of deafness. Most of these factors affect auditory speech recovery in adult CI users (Blamey et al., 2013), implying that crossmodal compensation also impacts CI outcomes (Campbell & Sharma, 2014; Heimler et al., 2014). The dynamic of such reorganisation is not well established, but it has been suggested that crossmodal reorganisation can be fast (Merabet et al., 2008) and supported by latent multimodal circuit (H. J. Lee et al., 2007). In the present study, we did not find any correlation between the visual bias and the personal characteristics of the patients (CI experience, age of implantation,…) but we know that multisensory integration evolves as long as patients are recovering auditory functions with a progressive increase in the implication of the visuo-auditory integrative temporal areas (Strelnikov et al., 2015).

In conclusion, our data represent clear evidence that, even after several months or years of recovery of auditory function, CI deaf patients remain strongly influenced by visual information when the auditory signal is too ambiguous and insufficient to allow a correct perceptual decision. Such crossmodal influence is observed in speech and in non-speech situations, such as face–voice interactions, which are crucial to social interaction. The clear visual impact is probably supported by a strengthening of the connectivity that occurs specifically during deafness between the face and voice cortical areas.

Acknowledgements

We thank the cochlear-implanted and normally hearing patients for their participation in this study, M.L. Laborde for her help in collecting the data, and C. Marlot for her help with the bibliography. This work was supported by ANR Plasmody (ANR-06-Neuro-021-04 to PBa and OD) and ArchiCore (ANR-14-CE13-0033-02 to PBa), DRCI Toulouse (Direction de la Recherche Clinique et de l'Innovation to MM), the recurring funding of the CNRS (to OD, MM and PBa), and by Fondation pour la Recherche Médicale (AJE2012-14) and BBSRC BB/I022287/1 to PBe.

Authors' Contributions

P. Barone, O. Deguine, and P. Belin planned and organised the experiment and wrote the article. L. Chambaudie, K. Strelnikov, and M. Marx conducted the experiment and analysed the data. B. Fraysse organised the experiment.

Additional Information

The authors declare no competing financial interests.

References

Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron, 48*(5), 859-872.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol Aging, 12*(1), 12-21.

Barone, P., & Deguine, O. (2011). Multisensory processing in cochlear implant listeners. In F. G. Zeng, R. Fay, & A. Popper (Eds.), *Springer Handbook of Auditory Research. Auditory Prostheses: Cochlear Implants and Beyond* (pp. 365-382). New-York: Springer-Verlag.

Bayard, C., Colin, C., & Leybaert, J. (2014). How is the McGurk effect modulated by Cued Speech in deaf and hearing adults? *Front Psychol, 5*, 416.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *Br J Psychol, 102*(4), 711-725.

Belin, P., Campanella, S., Ethofer, T., & Schweinberger, S. (2012). Audiovisual Integration in Speaker Identification *Integrating Face and Voice in Person Perception* (pp. 119-134): Springer New York.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci, 8*(3), 129-135.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., Burke, E. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiol Neurootol, 18*(1), 36-47.

Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *J Neurosci, 31*(36), 12906-12915.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn Sci, 11*(12), 535-543.

Campbell, J., & Sharma, A. (2014). Cross-modal re-organization in adults with early stage hearing loss. *PLoS One, 9*(2), e90594.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med, 19*(9), 1141-1164.

Champoux, F., Lepore, F., Gagne, J. P., & Theoret, H. (2009). Visual stimuli can impair auditory processing in cochlear implant users. *Neuropsychologia, 47*(1), 17-22.

Chen, L. C., Sandmann, P., Thorne, J. D., Bleichner, M. G., & Debener, S. (2016). Cross-Modal Functional Reorganization of Visual and Auditory Cortex in Adult Cochlear Implant Users Identified with fNIRS. *Neural Plast, 2016*, 4382656.

Coez, A., Zilbovicius, M., Ferrary, E., Bouccara, D., Mosnier, I., Ambert-Dahan, E. (2008). Cochlear implant benefits in deafness rehabilitation: PET study of temporal voice activations. *J Nucl Med, 49*(1), 60-67.

Collett, E., Marx, M., Gaillard, P., Roby, B., Fraysse, B., Deguine, O. (2016). Categorization of common sounds by cochlear implanted and normal hearing adults. *Hear Res, 335*, 207-219.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M. (2008). Audio-visual integration of emotion expression. *Brain Res, 1242*, 126-135.

de Gelder, B., & Vroomen, J. . (2000). he perception of emotions by ear and by eye. *Cognition & Emotion, 14*(3), 289-311.

Desai, S., Stickney, G., & Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *J Acoust Soc Am, 123*(1), 428-440.

Doucet, M. E., Bergeron, F., Lassonde, M., Ferron, P., & Lepore, F. (2006). Cross-modal reorganization and speech perception in cochlear implant users. *Brain, 129*(Pt 12), 3376-3383.

Ethofer, T., Bretscher, J., Wiethoff, S., Bisch, J., Schlipf, S., Wildgruber, D. (2013). Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. *Neuroimage, 76,* 45-56.

Foxton, J. M., Brown, A. C., Chambers, S., & Griffiths, T. D. (2004). Training improves acoustic pattern perception. *Curr Biol, 14*(4), 322-325.

Fu, Q. J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol, 5*(3), 253-260.

Fu, Q. J., Chinchilla, S., Nogaki, G., & Galvin, J. J., 3rd. (2005). Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *J Acoust Soc Am, 118*(3 Pt 1), 1711-1718.

Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia, 47*(13), 2967-2974.

Green, T., Faulkner, A., Rosen, S., & Macherey, O. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *Journal of the Acoustical Society of America, 118*(1), 375-385.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4*(6), 223-233.

Heimler, B., Weisz, N., & Collignon, O. (2014). Revisiting the adaptive and maladaptive effects of crossmodal plasticity. *Neuroscience, 283*, 44-63.

Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014a). Brain systems mediating voice identity processing in blind humans. *Hum Brain Mapp, 35*(9), 4607-4619.

Holig, C., Focker, J., Best, A., Roder, B., & Buchel, C. (2014b). Crossmodal plasticity in the fusiform gyrus of late blind individuals during voice recognition. *Neuroimage, 103*, 374-382.

Joassin, F., Pesenti, M., Maurage, P., Verreckt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex, 47*(3), 367-376.

Kovacic, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *J Acoust Soc Am, 126*(2), 762-775.

Kral, A., & O'Donoghue, G. M. (2010). Profound deafness in childhood. *N Engl J Med, 363*(15), 1438-1450.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22*(2), 948-955.

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neurosci, 11*, 36.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiol Aging, 27*(8), 1155-1163.

Lazard, D. S., Innes-Brown, H., & Barone, P. (2014). Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. *Hear Res, 307*, 136-143.

Lee, D. S., Lee, J. S., Oh, S. H., Kim, S. K., Kim, J. W., Chung, J. K. (2001). Cross-modal plasticity and cochlear implants. *Nature, 409*(6817), 149-150.

Lee, H. J., Giraud, A. L., Kang, E., Oh, S. H., Kang, H., Kim, C. S. (2007). Cortical activity at rest predicts cochlear implantation outcome. *Cereb Cortex, 17*(4), 909-917.

Mahoney, J. R., Li, P. C., Oh-Park, M., Verghese, J., & Holtzer, R. (2011). Multisensory integration across the senses in young and old adults. *Brain Res, 1426*, 43-53.

Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. *Ear Hear, 36*(2), 239-248.

Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hear Res, 275*(1-2), 120-129.

Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P. (2013). Gender categorization in cochlear implant users. *J Speech Lang Hear Res, 56*(5), 1389-1401.

McGurk, H., & Macdonald, J. (1976). Hearing Lips and Seeing Voices. *Nature, 264*(5588), 746-748.

Merabet, L. B., Hamilton, R., Schlaug, G., Swisher, J. D., Kiriakopoulos, E. T., Pitskel, N. B. (2008). Rapid and reversible recruitment of early visual cortex for touch. *PLoS One, 3*(8), e3046.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol Sci, 15*(2), 133-137.

Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends Amplif, 16*(2), 67-82.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage, 119*, 164-174.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex, 17*(5), 1147-1153.

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

Rouger, J., Lagleyre, S., Demonet, J. F., Fraysse, B., Deguine, O., & Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Hum Brain Mapp, 33*(8), 1929-1940.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci U S A, 104*(17), 7295-7300.

Sadato, N., Okada, T., Honda, M., Matsuki, K., Yoshida, M., Kashikura, K. (2005). Cross-modal integration and plastic changes revealed by lip movement, random-dot motion and sign languages in the hearing and deaf. *Cereb Cortex, 15*(8), 1113-1122.

Sadato, N., Yamada, H., Okada, T., Yoshida, M., Hasegawa, T., Matsuki, K. (2004). Age-dependent plasticity in the superior temporal sulcus in deaf humans: a functional MRI study. *BMC Neurosci, 5*, 56.

Schierholz, I., Finke, M., Schulte, S., Hauthal, N., Kantzke, C., Rach, S. (2015). Enhanced audio-visual interactions in the auditory cortex of elderly cochlear-implant users. *Hear Res, 328*, 133-147.

Schweinberger, S. R., Kloth, N., & Robertson, D. M. (2011). Hearing facial identities: brain correlates of face--voice integration in person identification. *Cortex, 47*(9), 1026-1037.

Song, J. J., Lee, H. J., Kang, H., Lee, D. S., Chang, S. O., & Oh, S. H. (2015). Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. *Brain Struct Funct, 220*(2), 1109-1125.

Stein, B. E., & Rowland, B. A. (2011). Organization and plasticity in multisensory integration: early and late experience affects its governing principles. *Prog Brain Res, 191*, 145-163.

Strelnikov, K., Marx, M., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2015). PET-imaging of brain plasticity after cochlear implantation. *Hear Res, 322*, 180-187.

Strelnikov, K., Rouger, J., Demonet, J. F., Lagleyre, S., Fraysse, B., Deguine, O. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain, 136*(Pt 12), 3682-3695.

Stropahl, M., Plotz, K., Schonfeld, R., Lenarz, T., Sandmann, P., Yovel, G. (2015). Cross-modal reorganization in cochlear implant users: Auditory cortex contributes to visual face processing. *Neuroimage, 121*, 159-170.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Tremblay, C., Champoux, F., Lepore, F., & Theoret, H. (2010). Audiovisual fusion and cochlear implant proficiency. *Restor Neurol Neurosci, 28*(2), 283-291.

Vachon, P., Voss, P., Lassonde, M., Leroux, J. M., Mensour, B., Beaudoin, G. (2013). Reorganization of the auditory, visual and multimodal areas in early deaf individuals. *Neuroscience, 245*, 50-60.

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *J Vis, 8*(9), 14 11-11.

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol, 4*(10), e326.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci*, *17*(3), 367-376.

Voss, P., Collignon, O., Lassonde, M., & Lepore, F. (2010). Adaptation to sensory loss. *Wiley Interdiscip Rev Cogn Sci, 1*(3), 308-328.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2013). Dissociating task difficulty from incongruence in face-voice emotion integration. *Front Hum Neurosci, 7,* 744.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *J Neurosci*, *34*(20), 6813-6821.

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends Cogn Sci, 17*(6), 263-271.

Table 1. Patient characteristics

Pati-ent	Age (years)	Gender	Aetiology of deafness	Deafness duration (years)	CI experience (years)	Speech comprehension score (in % of words)	Pure Tone Average, dB	VIx inc	VIx cong	Slope A	Slope AV, male face	Slope AV, female face
CI01	66	F	unknown	32	7.5	100	42	0,04	0,04	0,31	0,83	0,36
CI02	43	M	chronic otitis	2	18	90	27	-0,53	-0,23	0,53	0,03	0,04
CI03	53	M	unknown	15	1	65	37	-0,57	0,07	0,26	0,10	0,08
CI04	69	F	unknown	10	6	75	37	-0,02	0,07	0,43	0,79	0,29
CI05	49	M	unknown	8	8	90	33	-0,08	0,08	0,18	0,18	4,82
CI06	52	F	unknown	16	14	100	40	-0,05	-0,07	0,50	0,37	0,24
CI07	66	F	congenital	8	6	80	38	-0,05	-0,02	0,36	0,31	0,49
CI08	86	F	unknown	7	1	75	NA	-0,20	0,33	0,07	0,11	0,14
CI09	61	F	otospongiosis	24	1.5	90	30	-0,10	-0,17	0,33	0,13	0,12
CI10	86	M	meningitis	0.09	8.5	100	25	0,00	-0,04	0,43	0,28	0,27
CI11	46	F	otospongiosis	20	13	90	47	-1,00	0,85	0,08	NA	NA
CI12	67	F	unknown	5	7	90	NA	-0,01	-0,09	0,46	0,19	0,16
CI13	43	M	antibiotics	2	6	70	NA	-0,10	-0,02	0,33	0,28	0,24
CI14	77	F	unknown	12	9	90	37	-0,07	-0,07	0,45	0,59	0,45

NA for slopes is indicated for patient C11 who was completely driven by visual stimulation in the audiovisual (AV) conditions ignoring the auditory stimulation. As it is impossible to calculate the slope of a flat line, he was excluded from the analysis of slopes in AV performance. As for the auditory (A) condition, he was included in the analysis of this condition. NA non available due to the low fitting to a sigmoid function.

Figure legends

Figure 1: Results of the voice-gender categorisation task in auditory-only (A) and in audiovisual (AV) conditions.

a. Theoretical sigmoid curves in A and AV conditions with definition of criteria used to compute influence of visual facial stimuli on vocal processing. Representative AUC (area under the curve for femininity and above the curve for masculinity**)** is provided as an example for the condition with presentation of the male face. **b**. Psychophysical curves showing performances of the normal-hearing subjects (NHS), in percent of female response for each voice of the auditory continuum in the three conditions. In AV condition, the three curves are indistinguishable, demonstrating the absence of crossmodal interactions. **c.** Psychophysical curves showing performances of the NHS in the two-channel vocoding condition (NHS 2C), in the three conditions (A, V, and AV). As for the original voice condition (panel b), we did not observe a significant influence of the visual presentation on voice-gender categorisation. **d.** Psychophysical curves showing performances of the cochlear-implanted patients (CIP), in percent of female response for each voice of the auditory continuum in the three conditions. During the incongruent AV presentations, we observed a shift of the sigmoid function toward the gender carried by the face, revealing significant crossmodal interaction. AUC: area under the curve; M: masculinity, F: femininity.

Figure 2: Influence of visual presentation of a face on the auditory gender categorisation.

Error bars represent bootstrap confidence intervals ($p < 0.05$). The visual impact factor is significantly higher (more negative) in cochlear-implanted patients (CI) than in normal-hearing subjects (NHS) or in NHS in the two-channel vocoding condition (NHS 2C). This demonstrates the impact of deafness on crossmodal interaction.

Figure 1

Click here to download high resolution image

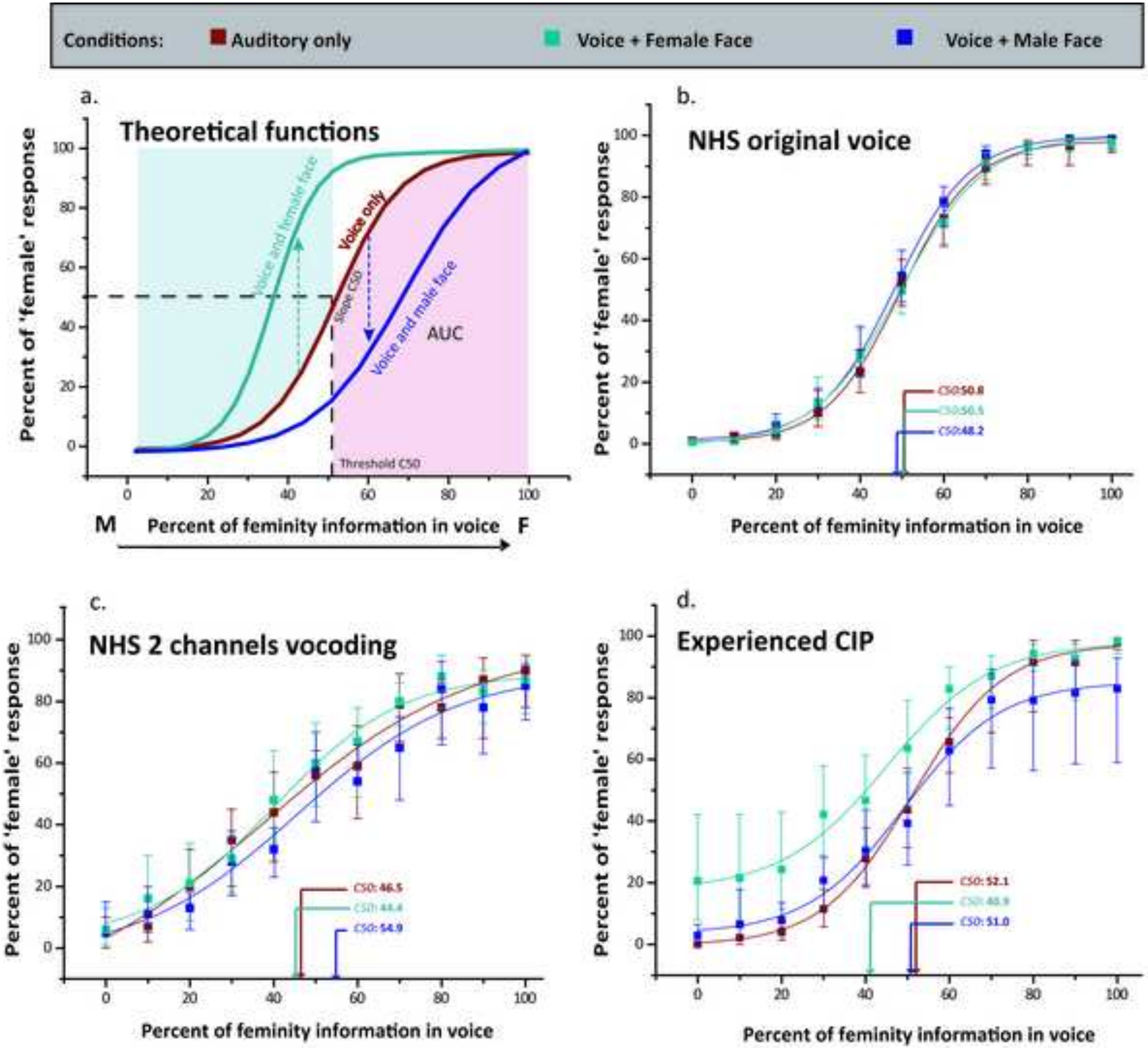
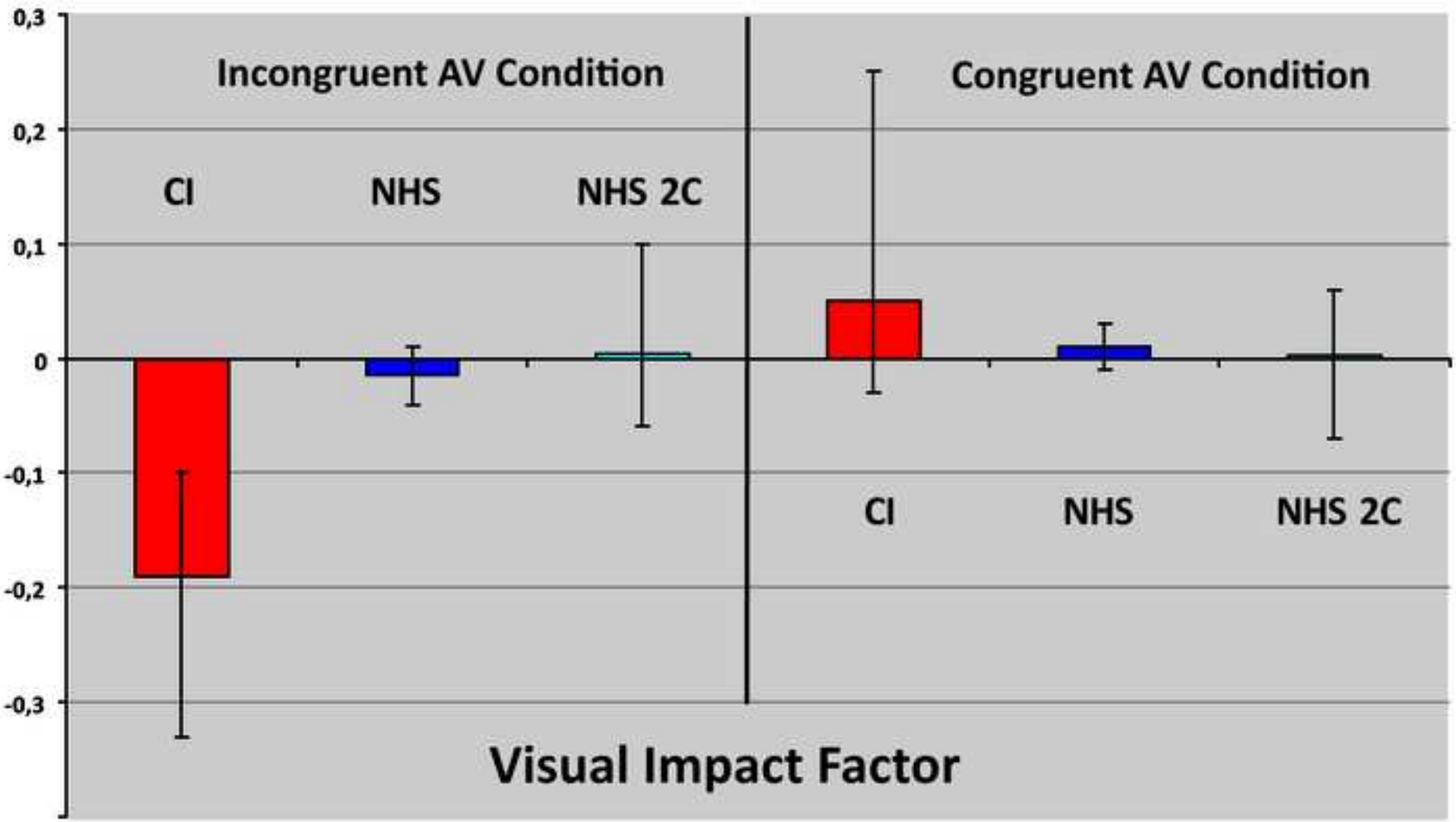


Figure 2


0,3
0,2
0,1
0
-0,1
-0,2
-0,3
Incongruent AV Condition
Congruent AV Condition
CI
NHS
NHS 2C
CI
NHS
NHS 2C
Visual Impact Factor