Processing of visuo-auditory prosodic information in cochlear-implanted patients deaf patients

Anne Lasfargues-Delannoy[1,2], *Kuzma Streilnikov*[1,2], *Sharon Ouddiz*[1], *Olivier Deguine*[1, 2], *Mathieu Marx*[1, 2], *Pascal Barone*[1]

[1]Cerveau & Cognition UMR 5549, Toulouse France
[2] Service d'Oto–Rhino–Laryngologie, Hôpital Purpan, Toulouse France
pascal.barone@cnrs.fr

Abstract

Linguistic prosody is a poorly treated subject in cochlear-implanted (CI) patients. Our study investigated how CI patients discriminate a question from a statement based only on linguistic prosodic cues in three conditions: visual, auditory, and visuo-auditory. The results demonstrate that CI patients are not better performers than normal-hearing subjects (NHS) in the visual only condition, but that they have better multi-sensory integration skills than controls. During audiovisual stimulation, CI patients fuse the auditory and visual information to enable a better discrimination of prosodic cues. This study confirms the importance of further research into CI patients prosodic discrimination skills notably concerning visual prosody and eye-tracking analysis.

Key words: audiovisual prosody, cochlear implant, multi-sensory gain

1. Introduction

Cochlear implantation (CI) is a rehabilitative neuroprosthesis destined for individuals with severe to profound bilateral hearing loss. This CI technology partially restores auditory function to enable speech comprehension, via an electrical stimulation of the auditory nerve [1]. Indeed, the auditory information given by the implant to the brain is degraded as it provides poor spectral information while preserving the temporal structures of sound [2]. Therefore the CI adequately restituates the sound envelope while the formants are spectrally and temporally degraded [3]. Hence, patients compensate using the visual sensory channel leading to audio-visual integration.

It is well known that speech is not just constructed from linguistic units (such as segmental units) but also from para-linguistic characteristics, such as prosody, which in turn depends on changes in frequency, amplitude, duration and rhythm. Prosody is often defined as the "*melody of speech"* [4], it is associated to the emotional content of the speech and from a linguistic point of view, it aids in distinguishing a statement from a question. Prosody conveys both linguistic and paralinguistic information and thereby plays a role in speech comprehension. Unfortunately, these prosodic structures are poorly coded by the implant due to the degraded spectral information [5]. This induces in cochlear implanted patients (CIP) difficulties in identifying gender, vocal identity, emotions carried by the voice as well as musical information [6] [7]. In CIP, the deficits present for the discrimination of the fundamental frequency [8] induces deficits in auditory speech prosody discrimination, especially if CIP do not have any residual hearing [5] [9]. Nonetheless, research has shown that visual information plays a crucial role, post cochlear implantation, in the recuperation of speech comprehension as patients use visual speech cues (i.e. lip reading) to aid their comprehension of auditory stimuli [3] [10] [11]. Since speech is a well-established multisensory stimulus, it is natural to assume that prosody is multisensory and has visual correlates. Indeed, there are visual correlates of prosody; certain facial movements are distinctly associated with auditory prosodic cues within a sentence [12]. For example: head movements are correlated with variations in the fundamental frequency of speech (pitch) and amplitude [13].The authors even suggest that the stressed word in a sentence can be identified solely on visual prosodic cues [13]. Further support for the presence of visual correlates of prosody emanate from Foxton et al., (2010) [14], who demonstrated that when participants are faced with congruence between visual and auditory prosodic cues, they perceive auditory prosodic cues as stronger. Hence, in bimodal stimulation, visual prosody significantly influences the perception of auditory prosody thereby facilitating cross modal perception of prosody. Furthermore, several studies have stated that when visual information is available the perception of prosody is facilitated [15] [13] [12].

1.1. Aims of the study

The aim of this study was to investigate how CI patients process visual prosodic cues compared to normal hearing subjects (NHS) based on the assumption that CI patients have developed visual compensatory skills making them better visual integrators for speech related information. Post-implantation, patients maintain supra-normal performances through time, in their speech reading skills compared to NHS, suggesting that the visual speech information counteract the degraded spectral information [16].Thereby suggesting the possibility of an adaptive synergy between visual and auditory information enabling better audio-visual gain than that observed in NHS. The research questions posited are the following:

- Do cochlear implanted patients benefit from auditory visual cues in the presence of prosodic ambiguity?
- Are CI patients better audio-visual integrators than NHS?

2. Materials and methods

2.1. Subjects

19 cochlear implanted (CI) patients and 30 normal hearing subjects (NHS) took part in this behavioural study on prosody.

All NHS were native French speakers with self-reported normal hearing and had self-reported normal or corrected-to-normal vision. CI patients were adults with a bilateral profound hearing loss acquired post-lingually due to diverse aetiologies. The duration of sensory deprivation varies according to each participant. The patients were either unilaterally or bilaterally implanted. Those implanted unilaterally had no residual hearing successfully compensated with hearing aids. At the time of this study all patients had more than 60% comprehension of dissyllabic words, indicating patients with good auditory recuperation and adaptation to the implant.
The age range of patients was comprised between 20 and 84 years with a mean of 57,8 ± 16,4 years old, 9 were women and 10 were male. Meanwhile the NHS group was composed of 20 women and 16 men with varying age ranges from 20-64 years of age with a mean age of 22,4 ± 2,6 years old.
All participants provided full informed consent prior to inclusion in this research protocol.

2.1.1. Preliminary data on old subjects with normal hearing

In order to counteract a possible age effect a second group of NHS was recruited. These 12 participants have a mean age of 57,9 ± 5,2 years (6 men and 6 women) which is a mean age similar to the mean age of the CI deaf patients. We performed an audiogram of this second group of subjects prior to inclusion. This subject group present normal or near to normal audiograms (according to audiology guidelines – between 0-20 dB) with a mean hearing loss of 19.3 ± 6.7 dB.

2.2. Stimuli

The stimuli presented to the participants were videos of unpunctuated spoken sentences, which could be posited to the subjects as either a question or a statement through the use of prosody. These sentences were allocated to three presentation conditions: visual only (V), auditory only (A) and audio-visual (AV).

The videos were recorded with a professional digital camera. Each sentence was then isolated from the main recording using window moviemaker® and the final stimuli was created. In order to diminish the language component of the task the unpunctuated written sentence was presented prior to the video.

2.3. Strategies inherent to this study

In order to better compare the data obtained from CI patients and NHS, we controlled certain parameters in the creation of the stimuli. Firstly we asked the two French actors to reduce their facial expressions to diminish the natural visual cues associated with a question or a statement, thereby, increasing the difficulty in distinguishing these two linguistic items. This was also done to avoid a ceiling effect in the performances of CI patient as previous research has shown enhanced visual capacities in this patient population due to the crossmodal compensation mechanisms.
Secondly, for the NHS, we degraded the auditory information using a vocoding system, of 8 canals, for the auditory and audio-visual conditions. The vocoding simulates the process that occurs in a cochlear implant as the temporal information of the sound is preserved but the fine acoustical structures are degraded [3]. The use of vocoding avoids the ceiling effect obtained if NHS are presented with normal auditory stimuli. Furthermore controlling these parameters, through the use of these two strategies, will permit a better comparison between CI patients and normal hearing controls when comparing each modality.

2.4. Stimuli presentation

For each condition, the participants are presented with a fixation cross, followed by a written sentence (without any punctuation) then another fixation cross and the video in accordance with the condition presented (V, A, VA). At the end of the video, the subjects have to make a judgement on the prosodic form of the sentence: was it a question or a statement?
The stimuli were presented through Matlab®, on a portable PC, speakers were located beside the computer screen and the volume was calibrated to be between 55 and 65dB. The order of the stimuli presentation, within each condition was random. The visual condition was presented first for each subject, followed by the audio-visual and auditory conditions, which varied alternatively in their presentation order.

2.5. Data analysis

The behavioural data was collected in terms of raw performance, i.e. the percentage of correct responses, and analysed using excel®, matlab® and R®. Regarding each condition presented to the participants, a comparative analysis was carried out between the two subject groups. For this a mixed linear effect model was used to compare the percentage of correct responses. Additionally, the raw data was also transformed into d'prime values, in order to obtain the sensibility value and counteract the decisional bias of each individual. Furthermore, we calculated the visuo-auditory gain of CI patients to the one obtained by NHS. The VA gain was calculated individually by normalizing the performances in the VA condition with respect to the best unimodality (either visual or auditory). This was done to allow a fair comparison of the VA gain. The statistical analysis of the VA gain was calculated using a bootstrap, as this data is non-parametric.

2.6. Further research

In addition to the behavioural analysis, we investigated, through eye tracking analysis using the Tobii X2-60® system, the facial exploration mechanisms used by the participants during the visual and visuo-auditory condition, to identify differences in the exploration strategies between these two V and AV conditions when discriminating a question from a statement.

3. Results

3.1. Behavioural analysis across conditions and groups

Based on our strategy of minimizing the visual information (for both NHS and CIP) and degrading the auditory information (for NHS) the behavioural data retrieved from CI patients and NHS leads to an absence of statistical difference between the two groups concerning their unimodal performances, according to our mixed effect linear model (visual condition: t-value=0.4748, p= 0.0821; auditory condition: t-value=0.4748, p= 0.667; visuo-auditory condition: t-value=0.4748, p= 0.717). Therefore our two groups of

subjects have similar performances and this enables a more robust comparison of the visuo-auditory gain. In addition, our results clearly show a non-significant (p= 0.667) tendency, for lower performance rate, in CIP, for the visual condition (69%) compared with NHS (79%) – see figure 1.

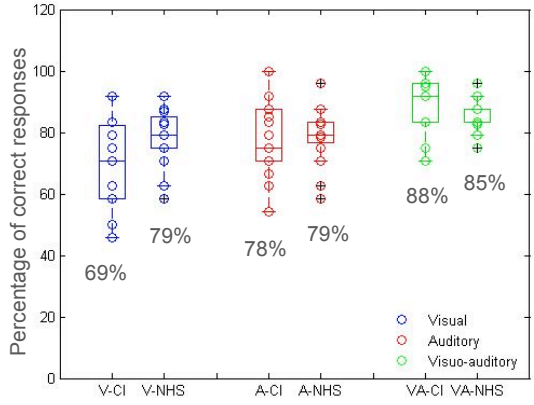

Figure 1: *CI patients and NHS performances with data spread and mean performance for each condition.*

This indicates that, contrary to our previous hypothesis, CI patients do not necessarily have better visual capacities concerning the identification of visual prosodic cues. In addition, we observed that 74% of CIP were better performers in the auditory condition than in the visual condition (26%). In NHS only 60% are better performers in auditory discrimination vs. 40% for the visual condition. However, this could simply be the outcome of better adaptation from CIP to "distorted" speech compared to NHS who are "naïve" listeners with respect to vocoded sounds.

Nonetheless, we observe a strong variability in CI patients for the unimodal conditions, due to the large spread of data observed (Figure 1). A variance analysis carried out reveals a significant variability in CIP ($p<0.05$) not observed in NHS. This variance observed in CIP, in the V only condition, reduces significantly when compared to the variability obtained in the bimodal (VA) condition. This significant difference suggests that CIP benefit from multisensory integration when faced with a bimodal stimulus. This enables a significant reduction in their variability.

3.2. The visual auditory gain: a fusion between vision and hearing

In both groups, the bimodal presentation induced an increase in performances (Figure 1) in agreement with the well-known perceptual benefits of multisensory integration. As the unimodal performances were similar in both groups, calculating the VA gain allows to us to observe if NHS and CIP differ with respect to bimodal integration for prosodic stimuli. Figure 2 shows VA gain values for CI patients and NHS. The bootstrap analysis carried out on this data reveals a significant difference within the CIP group and between CIP and NHS as the confidence intervals do not overlap between these two groups. CIP have a MS gain, which is about 4 times greater to that observed in NHS (13.4% vs. 3.2%). Of importance, the MS gain observed in NHS does not reach significance. Altogether these results highly suggest the presence of a strong skill in CIP to integrate visual and auditory prosodic information.

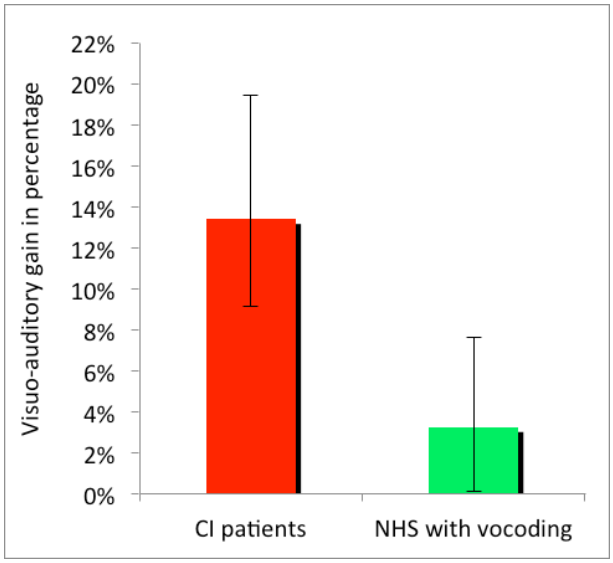

Figure 2: *The VA gain for CI patients and NHS with confidence interval values.*

3.3. Preliminary data on old subjects with normal hearing

We recruited a group of aged subjects with a mean age similar to the one of the CIP group. These subjects present hearing threshold, which are considered as reflecting normal hearing capacity on the speech conversational frequencies. The performance level of these aged subjects is similar to that observed in younger NHS and in CI patients as evidenced by table 1. There are no statistical differences between the old and young NHS groups irrelevant of the condition tested according to the mixed effect linear model. Of importance, older NHS present a weak VA gain of 3.7% which is not different to that obtained in younger hearing subjects ($p>0.05$) and significantly lower to the MS benefits presented by CI patients. As for the younger group, the inferior and upper confidence intervals were not different from zero a result that suggest that this VA gain is not statistically significant. In conclusion, these preliminary data favor better multisensory integration of auditory and visual stimuli in CI patients.

	V condition	A condition	VA condition	VA gain
Young NHS	79.4%	78.6%	84.8%	3.2%
Old NHS	77.5%	71.5%	81.7%	3.7%
CIP	68.6%	77.5%	88.5%	13.4%

Table 1: Raw performance data for CI patients, Young NHS and Old NHS.

3.4. Preliminary eye tracking data

Since CI patients did not demonstrate better visual capacities to aid in the discrimination between a question and a statement based on visual prosodic cues, this study was prolonged with eye-tracking data, in order to see how CI patients and NHS explore the face to aid in this distinction.

For this preliminary data, 10 NHS were tested and the total number of fixations in the different visual regions of interest was calculated. Whilst comparing the visual and the visuo-auditory condition we can observe a slightly more important tendency for NHS to fixate on the eyes 22,4% during the visual condition vs. 18.28% for the visuo-auditory condition. However this result is not currently significant. The heat map below shows the difference between the fixation patterns of 2 CI patients and NHS for the visual only condition for both a question and a statement.

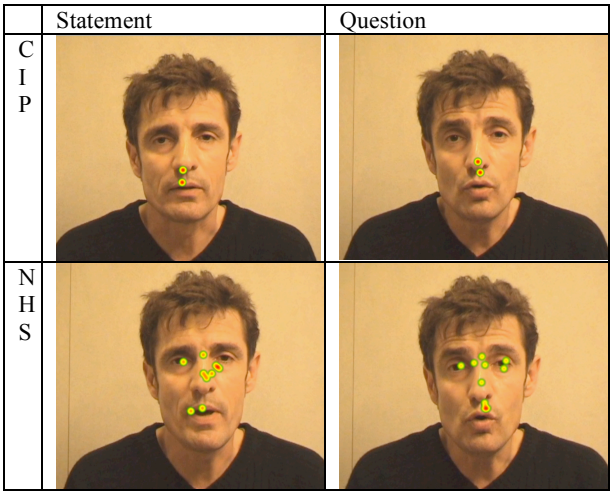

Figure 3: *Fixation patterns for two CI patients and NHS during the visual condition*

The two CI patients currently recruited tend to show a clear preference for the lower half of the face. These results suggest that CIP are automatically engaged in a lip-reading task in spite of the fact that they have been provided with the written sentence before the video presentation, and that sentence comprehension was not an issue to perform the task. However, more data is necessary to confirm that CIP present an abnormal face exploration during prosodic cue discrimination.

4. Discussion

4.1. Multisensory integration in speech

It is a widely known phenomenon, that the visual sensory channel enhances auditory speech perception, in a noisy environment as a result of multisensory integration (MSI); the visual and auditory information are complementary and can be fused to enable better speech discrimination in noise. CI patients also rely on MSI to enhance speech perception to disambiguate the degraded auditory information brought by the implant. Rouger et al. (2007) [3], demonstrated that, post cochlear implantation, patients rapidly obtain good speech comprehension in bimodal stimulation (between 80% and 100%) compared to unimodal performances. Concerning, linguistic prosodic cues, we proposed that a similar phenomenon can occur in CIP. In accordance with our hypothesis, visual and auditory prosodic cues are complementary and when presented together enable a VA gain, as evidenced by our results. Indeed, the presence of a bimodal stimulation inherently reduces the ambiguity that a unimodal stimulus can contain, due to the presence of redundant prosodic cues. Our results support the presence of a multisensory integration in CI patients as we observed significantly better performances in the bimodal condition and a significant MS gain of 13% suggesting a cross-modal interaction between visual and auditory prosodic cues. On the opposite, in NHS such benefit is weak and non-significant that could be explain by a ceiling effect. According to the inverted effectiveness rule, we expect to have higher MS gain in NHS in more difficult unimodal conditions.

4.2. CI patients and visual cues

It is often assumed that CI patients have supra-normal visual capacities due to their impaired auditory skills. However, CI patients, contrary to prior beliefs, do not have supranormal visual prosodic discrimination skills in our experimental difficult conditions. Indeed, our results show that they even have slightly lower (non- significant) performances in the visual unimodal condition than NHS (69% vs 79%). The primary hypothesis of such result is that distinguishing a statement from a question is not a meaningful and logic task for CI patients. To distinguish such differences in their everyday life they probably rely mainly on grammatical and syntactic cues. Secondly, it could be hypothesised that CI patients have different oculomotor behaviours when facing a speaking person. We hypothesize that CIP present a systematic bias of face exploration towards the lower half of the face, due to their enhanced and probably automatic, speech-reading skills. Such behaviour, present in our restricted set of patients, may harm CI patients in the treatment of visual prosodic cues because these cues tend to be associated with eyebrow movements, head movements etc.. As CI patients focalise their attention on the lips they may not perceive these visual prosodic cues that aid in identifying a question from a statement. Research on congenitally deaf individuals that communicate through sign language have shown, through eye tracking analysis, that these patients tend to fixate the nose and mouth region, and have better peripheral vision to discriminate sign language information [17]–[19]. Such skill is probably absent in adult CI deaf patients.

4.3. CI patients have a higher multisensory gain

The multisensory gain, or VA gain, observed for speech perception [3] and prosodic discrimination due to the redundant information, is interpreted as a reduction of the ambiguity that a unimodal stimulus can contain. Our results support the presence of a significantly higher multisensory integration in CI patients than NHS. Of interest, our preliminary analysis on a group of aged NHS confirms that the increased AV gain observed in CIP is not the consequence of a different multisensory mechanism resulting from aging processes. Indeed, old and young NHS present a similar weak multisensory gain, a result that could suggest that aging has no impact on the capacity of integrating auditory and visual prosodic cues, at least concerning this age range (50-65 years). Lastly, the fact that CIP present a much higher VA gain compared to age-matched NHS, suggests that this multisensory benefits does not originate from an adaptive strategy that could be developed specifically in aged subjects, but rather that it is the consequence of an adaptive strategy induced by deafness and cochlear implantation. These results confirmed that CI patients, independently of their age are more proficient for the fusion of visual and auditory linguistic and paralinguistic information. During the prolonged period of deafness CI patients have developed a compensatory strategy based on the visual channel. After cochlear implantation CI

patients rely more on visual information to compensate for the distorted/impoverished auditory information provided by the implant. We postulate that CI patients are better multisensory integrators for prosodic information because this A-V fusion appears to be independent of their visual skills as, in our experimental protocol, they are not better than NHS during the visual only discrimination tasks. These results complement those obtained by Rouger et al. (2007) [3] regarding speech comprehension. Post cochlear implantation, patients rapidly obtain better AV speech comprehension performances compared to NHS evaluated with a simulation of a cochlear implant. CI patients are therefore superior cross modal integrators since they use the redundant prosodic and speech cues to better treat bimodal stimuli. This supra-normal multisensory integration capacity that CI patients appear to have is dependent on a strong audio-visual synergy, which has previously been demonstrated at the brain level [20], [21].

5. Conclusions

Linguistic prosody is not only an auditory stimulus but also a visual one. NHS and CI patients both treat visual prosodic cues, but contrary to visual speech (lip reading) CI patients are not better at identifying these visual prosodic indices than NHS. Nevertheless, the auditory prosodic cues, which are inadequately rendered by the cochlear implant, induce a deficit in CI patients capacity to correctly distinguish a question from a statement based only on auditory capacities. However, this perceptual deficit of auditory prosody is well compensated by CI patients in audio-visual stimulation due to multisensory integration. Indeed, CI patients demonstrate significantly better bimodal integration than NH counterparts. This suggests the presence of enhanced synergy between auditory and visual cues in CI patients, as a compensation mechanism.

Regarding the visual exploration of the face, the data presented here is preliminary data that require caution in its interpretation as well as the pursuit in this line of enquiry. Undeniably, NHS do not have a different exploration strategy irrelevant of the test condition. We await further CI patient data to enable a comparison with NHS, and statistical analyses.

6. Acknowledgements

We would like to thank the CI patients and NH subjects for their participation, ML. Laborde, C. Algans and M. Tartayre for their help in the recruitment process, and the ENT service for their support.
This work was supported by ANR Archicore (ANR-14-CE13-0033-02) and ANR VirtualHearing3D (ANR-16-CE17-0016-03), and recurrent funding of the CNRS.

7. References

[1] B. S. Wilson, C. C. Finley, and D. T. Lawson, “Cochlear Implants: Models of the Electrically Stimulated Ear,” 1990.

[2] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, “Better speech recognition with cochlear implants,” *Nature*, vol. 352, no. 6332, pp. 236–238, Jul. 1991.

[3] J. Rouger, S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, and P. Barone, “Evidence that cochlear-implanted deaf patients are better multisensory integrators.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 17, pp. 7295–300, 2007.

[4] A. Wennerstrom, *The music of everyday speech : Prosody and discours analysis.* .

[5] P. Barone *et al.*, “Speech Prosody Perception in Cochlear Implant Users With and Without Residual Hearing Speech Prosody Perception in Cochlear Implant Users With and Without Residual Hearing,” no. August 2015, pp. 239–248, 2014.

[6] P. Belin, S. Fecteau, and C. B??dard, “Thinking the voice: Neural correlates of voice perception,” *Trends Cogn. Sci.*, vol. 8, no. 3, pp. 129–135, 2004.

[7] Z. Massida *et al.*, “Voice discrimination in cochlear-implanted deaf subjects,” *Hear. Res.*, vol. 275, no. 1–2, pp. 120–129, 2011.

[8] S. C. Peng, N. Lu, and M. Chatterjee, “Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners,” *Audiol. Neurotol.*, vol. 14, no. 5, pp. 327–337, 2009.

[9] M. Marx *et al.*, “Speech Prosody Perception in Cochlear Implant Users With and Without Residual Hearing,” *Ear Hear.*, vol. 36, no. 2, 2015.

[10] D. S. Lazard *et al.*, “Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time.,” *PloS One*, vol. 7, no. 11, p. e48739, Jan. 2012.

[11] C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, and B. C. J. Moore, “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,” *Proc. Natl. Acad. Sci.*, vol. 103, no. 49, pp. 18866–18869, Dec. 2006.

[12] E. Cvejic, J. Kim, and C. Davis, “Prosody off the top of the head: Prosodic contrasts can be discriminated by head motion,” *Speech Commun.*, vol. 52, no. 6, pp. 555–564, 2010.

[13] K. G. Munhall, J. a Jones, D. E. Callan, T. Kuratate, and E. Vatikiotis-Bateson, “Visual prosody and speech intelligibility: head movement improves auditory speech perception.,” *Psychol. Sci. J. Am. Psychol. Soc. APS*, vol. 15, no. 2, pp. 133–137, 2004.

[14] J. M. Foxton, L. D. Riviere, and P. Barone, “Cross-modal facilitation in speech prosody,” *Cognition*, vol. 115, no. 1, pp. 71–78, 2010.

[15] H. C. Yehia, T. Kuratate, and E. Vatikiotis-Bateson, “Linking facial animation, head motion and speech acoustics,” *J. Phon.*, vol. 30, no. 3, pp. 555–568, 2002.

[16] P. Barone, K. Strelnikov, and O. Deguine, “Auditory recovery and speechreading in cochlear implanted deaf patients : a review,” *Audiol Med*, vol. 8, no. 2, pp. 100–106, 2010.

[17] D. Bavelier, M. W. G. Dye, and P. C. Hauser, “Do deaf individuals see better?,” *Trends Cogn. Sci.*, vol. 10, no. 11, pp. 512–518, 2006.

[18] D. Bavelier *et al.*, “Visual attention to the periphery is enhanced in congenitally deaf individuals.,” *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 20, no. 17, p. RC93, 2000.

[19] C. L. De Filippo and C. R. Lansing, “Eye fixations of deaf and hearing observers in simultaneous communication perception.,” *Ear Hear.*, vol. 27, pp. 331–352, 2006.

[20] K. Strelnikov, J. Rouger, P. Barone, and O. Deguine, “Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants,” *Scand. J. Psychol.*, vol. 50, no. 5, pp. 437–444, 2009.

[21] K. Strelnikov *et al.*, “Increased audiovisual integration in cochlear-implanted deaf patients: Independent components analysis of longitudinal positron emission tomography data,” *Eur. J. Neurosci.*, vol. 41, no. 5, pp. 677–685, 2015.