
***This is a self-archiving copy and does not fully replicate the published version***

Audio-visual speech perception in noise: Implanted children and young adults versus normal hearing peers

Riki Taitelbaum-Swead\textsuperscript{a,b}, Leah Fostick\textsuperscript{a}

\textsuperscript{a} Ariel University, Department of Communication Disorders
\textsuperscript{b} Meuhedet Health Services

**Corresponding Author:**

Riki Taitelbaum-Swead (PhD)
Ariel University, Department of Communication Disorders
Kiryat HaMada 3, Ariel, Israel 40700
Tel: 972-3-9765755; Fax: 972-3-9758908; E-mail: rikits@ariel.ac.il

Leah Fostick (PhD)
Ariel University, Department of Communication Disorders
Kiryat HaMada 3, Ariel, Israel 40700
Tel. 972-3-9765755; Fax. 972-3-9758908; E-mail: leahfo@ariel.ac.il
ABSTRACT

Objective: The purpose of the current study was to evaluate auditory, visual and audiovisual speech perception abilities among two groups of cochlear implant (CI) users: prelingual children and long-term young adults, as compared to their normal hearing (NH) peers.

Methods: Prospective cohort study that included 50 participants, divided into two groups of CI (10 children and 10 adults), and two groups of normal hearing peers (15 participants each). Speech stimuli included monosyllabic meaningful and nonsense words in a signal to noise ratio of 0 dB. Speech stimuli were introduced via auditory, visual and audiovisual modalities.

Results: (1) CI children and adults show lower speech perception accuracy with background noise in audiovisual and auditory modalities, as compared to NH peers, but significantly higher visual speech perception scores. (2) CI children are superior to CI adults in speech perception in noise via auditory modality, but inferior in the visual one. Both CI children and CI adults had similar audiovisual integration.

Conclusions: The findings of the current study show that in spite of the fact that the CI children were implanted bilaterally, at a very young age, and using advanced technology, they still have difficulties in perceiving speech in adverse listening conditions even when adding the visual modality. This suggests that adding
audiovisual training might be beneficial for this group by improving their audiovisual integration in difficult listening situations.

**Key words**: cochlear implant, audio visual speech perception, children, speech reading

1.1 INTRODUCTION

Cochlear implants (CI) are now the standard of care for hearing rehabilitation of severe to profound hearing loss in children and adults [1-2]. A large number of studies have shown the success of cochlear implants in providing better sound accessibility and enabling better speech perception and production, among both children and adults [3-6]. This success, and the progress of technology, has resulted in an expansion in the candidacy criteria for CI, including a decrease in age at implantation (resulting in an increasing number of infants being implanted in the second half of their first year of life [7]), and more cases of bilateral implants [8]. The combination of the availability of good hearing devices available at a very young age, and, when needed, to both ears, produces improved accessibility of auditory information [3, 9]. Some studies have even shown that many implanted children can achieve age-appropriate levels of speech perception, similar to their normal hearing peers, when tested under optimal conditions (e.g. with no background noise) [7]. However, under adverse listening conditions, such as the presence of noise, the performance of cochlear implanted individuals is still reduced [10-11].

The enormous range of environmental listening conditions during daily life communication places huge demands on the process of speech perception. In transit from speaker to listener, speech signals are often altered by background noise and
other interfering signals. The performance of CI users under these conditions has been found to be deficient [5]. This deficiency is especially significant, since many implanted children are educated in mainstream settings where the signal to noise ratios (SNR's) in classrooms may be as high as -6 dB [12]. Such conditions can make it very difficult for CI users to concentrate and learn new academic material.

Previous studies have tried to quantify the performance of prelingual CI users under different listening conditions [5,6]. Most of these studies, however, focused on examining speech perception via the auditory modality. Yet, under natural conditions, the CI user often experiences face to face situations in which visual information is also available. Indeed, speech perception accuracy have been demonstrated to improve when speech reading is added to auditory input both among the normal hearing population [13-15] and CI users [16-18, 19]. However, only few studies tested both CI and NH participants on audiovisual speech perception. One study used under background noise condition, but only for the normal hearing participants (to avoid ceiling effect) [16]. Therefore, no group comparison was made. Another study used under different presentation levels (speech detection and recognition thresholds) for CI and NH participants [19], which clearly differ between groups. Identifying potential differences in audiovisual speech perception among these two populations would help shed light on how prelingual CI users employ each modality in perceiving speech, compared to NH participants. This was the aim of the current study.

When studying perceptual achievements of CI users, age has important implications. First, age of implantation has a huge effect on speech perception: those who were implanted at an earlier age perform better than those implanted at older age [4]. Second, the current age of the CI user needs to be considered. Guidelines for
implantation have changed over the years with regard to the minimum age of implantation, amount of residual hearing required, and the option of bilateral implantation. These factors contribute to making the population of less recent implantees (long term CI users who are now young adults) inherently different from the more recent ones (children). A third implication of age is the sensory system maturation that occurs with time. We and others [13, 20-23] have previously shown that each modality develops at a different rate; while auditory and audiovisual speech perception matures already by the age of 8-9, visual speech perception matures only after this age. Therefore, in the present study, we compare auditory, visual, and audiovisual speech perception between prelingual CI children and long-term CI user adults, and their NH peers, in order to determine how CI affect auditory, visual and audiovisual speech perception in different age groups.
1.2 MATERIAL AND METHODS

1.2.1 Participants

The study included 50 individuals, divided into two groups of CI (children and long term young adult users), and two groups of normal hearing individuals. The adults were matched by age and the children by hearing age. All participants possessed Hebrew as a native language, normal or corrected visual ability, no reported developmental, cognitive or neurological problems, and normal speech and language abilities (based on parental report, in the case of children).

1.2.1.1 CI groups

The CI groups included 20 implanted participants: 10 children (CI children) and 10 young adults (CI adults). All implanted participants met the following inclusion criteria: (1) Onset of severe to profound hearing impairment before age 3; (2) Usage of hearing aids prior to implantation; (3) Mainstream education and oral communication; (4) At least 50% in monosyllabic open set test in quiet; (5) CI devices of Cochlear, Advanced Bionics, or Med-El.

The CI children’s mean chronological age was 6.5 years (S.D. 0.9 years). All children used two implants. The mean age at first implantation was 16.3 months (S.D. 5.6 months), and the mean age of second implantation was 32.6 months (S.D. 19.6 months). Etiology of the hearing loss was genetic for seven of the CI children and unknown for three. The CI young adults’ mean chronological age was 22.6 years (S.D. 2.05 years). Their mean age at first implantation was 9.1 years (S.D. 7.0); five of the participants used two implants, three participants used one implant with no hearing aid in the second ear, and two participants used an implant on one side and a hearing aid on the other side. The mean age at second implantation was 13.9 (S.D
7.1). Etiology of hearing loss was genetic for six of the CI adults and unknown for four.

1.2.1.2 Normal hearing groups

Normal hearing participants included 15 children aged 4-5 years and 15 young adults aged 20-30 years. All participants had normal hearing thresholds (pure-tone air-conduction thresholds less than 15 dB HL bilaterally at octave frequencies from 250 - 4,000 Hz [24]).

1.2.2 Speech perception tests

Speech perception tests were monosyllabic meaningful and nonsense words, which include mainly acoustic information and minimal linguistic redundancy.

1.2.2.1 Meaningful Words

Monosyllabic meaningful Hebrew AB lists (based on [25]) were used in the present study. This test includes 12 lists (narrated in a film, as described in the sub-section on Apparatus below), each consisting of ten monosyllabic words. Each list contains ten syllables in a consonant-vowel-consonant (CVC) pattern in which the 19 consonants of the Hebrew language appear either at the initial or the final position, and each of the five Hebrew vowels (/a/, /e/, /i/, /o/, /u/) appears twice.

1.2.2.2 Nonsense Words

This test resembles the structure of the meaningful words test and also includes 12 lists of ten monosyllabic CVC syllables. However, in this test, the syllables are nonsensical but contain some phonological redundancy, in accord with Hebrew
linguistic rules (for example, the consonants /b/ and /p/ never appear in the final position).

1.2.3 Apparatus

A female native Hebrew speaker with intelligible articulation and clear facial movements was filmed and recorded. The speaker looked directly into the camera, starting and ending each utterance with a natural face/closed mouth position. The speaker was recorded against a bright background in a quiet, well-lit recording studio. Her face appeared in full on the entire screen. The audiovisual recordings were digitized using Apple Final Cut Pro X software with 64-bit resolution.

The words were recorded in a studio using a SONTRONICS TCS-6 microphone and Samplitude classic 8.1 recording software. They were edited using the Sound Forge program, which digitized (16-bit) at a sampling rate of 44 kHz. Word level was normalized using the overall Root Mean Square (RMS). White noise generated by the Sound Forge program was added to the normalized words in a signal-to-noise ratio (SNR) of 0 dB. The noise was added to words in the audiovisual and auditory conditions, while in the visual condition the words were presented in quiet.

The words were presented using the Winamp Media Player 5.7 software, via S-Tech supra aural headphones for the NH groups, or via sound field for the CI groups. Words intensity was 70 dBSPL, as measured by a TA 1350A Sound Level Meter were used for the NH groups. Participant responses were recorded using a SONY ICD-PX312 recording device placed in close proximity to the participants. An inter-acoustic AD229B audiometer was used to screen and measure hearing levels.
1.2.4 Procedure

The study was approved by the Institutional Review Board and was conducted in accordance with Good Clinical Practice (GCP) guidelines. All participants received a full explanation about the study and provided signed informed consent. For participants under age 18, parental informed consent was obtained and assent to participate was obtained from the children. All potential participants’ in the NH groups were screened for hearing levels prior to participation in the study.

The study included six different conditions: two context conditions (meaningful and nonsense words) and three modality conditions (Auditory, Visual, or Audiovisual). Two lists (20 words) were presented for each condition. The order of the words within each list, as well as the order of meaningful and nonsense words, were randomly ordered for each participant. Each participant was tested under all six conditions.

Participants were instructed to repeat each word immediately after hearing it or after viewing the video, and their responses were recorded. Two independent raters transcribed these responses. When the raters disagreed (this occurred in less than 1% of the words), a transcription of a third rater was used.

The study was not supported by any sponsor or funding agency.

1.2.5 Statistical analysis

The main analysis for the effect of type of word, modality, and group on speech perception accuracy was carried out using a two-way repeated measures Analysis of Variance (ANOVA). Type of word (meaningful, nonsense) and modality (audiovisual, audio, visual) were entered as within-subjects variables, and group (CI children, CI adults, NH children, NH adults) as a between-subjects variable. Simple
effects were analyzed using one-way ANOVA and t-tests. Post-hoc analyses were done using Least Significant Differences (LSD).
1.3 RESULTS

Table 1 presents means and standard deviations (SDs) of the study groups’ speech perception accuracy for meaningful and nonsense words across each modality. Significant main effects were found for group \( (F_{(3,48)} = 49.955, p=.000) \), modality \( (F_{(2,96)} = 444.475, p=.000) \), and type of word \( (F_{(1,48)} = 68.776, p=.000) \). As expected, meaningful words (mean=39.28, SD=7.69) obtained higher accuracy rate than nonsense words (mean=29.390, SD=8.93).

| INSERT TABLE 1 |

Post-hoc analysis showed that, in general, speech perception accuracy for CI children was not different from CI adults’ (LSD=.833, p=.789), but was significantly lower than NH children’s (LSD=-16.639, p=.000) and NH adults’ (LSD=-27.201, p=.000). CI adults had also lower speech perception accuracy than both NH children (LSD=-17.472, p=.000) and NH adults (LSD=-28.034, p=.000). NH children had lower speech perception accuracy than NH adults (LSD=-10.562, p=.000). The audiovisual modality produced an overall higher accuracy rate than both the auditory (LSD=19.192, p=.000) and visual modality separately (LSD=48.214, p=.000). The auditory modality produced an overall higher accuracy rate than the visual (LSD=-29.022, p=.000).

Significant modality by group and modality by type of word interactions were observed \( (F_{(6,96)}=53.801, p=.000 \text{ and } F_{(2,96)}=9.429, p=.000, \text{ respectively}) \). Three separate one-way ANOVAs were carried out for each modality on accuracy data, averaged for meaningful and nonsense words (Table 1). Significant group effects were found for all modalities (audiovisual: \( F_{(3,48)}=43.705, p=.000 \); auditory: \( F_{(3,48)}=65.332, p=.000 \); visual: \( F_{(3,48)}=22.281, p=.000 \)).
Figure 1 presents speech perception accuracy, averaged for meaningful and nonsense words, for all study groups across each modality. Post-hoc LSD analyses revealed that CI children were not different in speech perception accuracy from CI adults in the audiovisual modality (LSD=2.000, p=.687), but had a higher accuracy rate than CI adults in the auditory modality (LSD=11.750, p=.024) and lower accuracy than CI adults in the visual modality (LSD=-11.250, p=.000). CI children had lower accuracy than NH children in both audiovisual and auditory modalities (LSD=-25.333, p=.000 and LSD=-14.363, p=.001, respectively), but had a higher accuracy rate than the NH children in the visual modality (LSD=4.833, p=.023). Similarly, CI adults had lower accuracy than NH adults in both audiovisual and auditory modalities (LSD=-41.882, p=.000 and LSD=55.529, p=.000, respectively), and higher accuracy than the NH children and adults in the visual modality (LSD=-13.309, p=.000 and LSD=13.308, p=.000, respectively). As expected, NH children had lower accuracy than NH adults in both audiovisual and auditory modalities (LSD=-14.549, p=.001 and LSD=-14.363, p=.001, respectively), but were not different from NH adults in the visual modality (LSD=-2.775, p=.126). No difference was found between CI children and NH adults in the visual modality (LSD=2.058, p=.309).

Figure 2 presents speech perception accuracy for meaningful and nonsense words in each modality, averaged across study groups. Three separate t-tests were performed on meaningful and nonsense words across each modality. Significant effects for type of word was found for all of the modalities (audiovisual: $t_{(51)}=6.107$, $p=.000$; auditory: $t_{(51)}=6.470$, $p=.000$; visual: $t_{(51)}=2.675$, $p=.010$), however, the difference in accuracy between meaningful and nonsense words was smaller in the
visual modality than in the audiovisual and auditory modalities. No group by type of word or modality by group by type of word interactions were observed ($F_{(3,48)}=2.220$, $p=.098$ and $F_{(6,96)}=.653$, $p=.688$, respectively).

INSERT FIGURE 2
1.4 DISCUSSION

The present study shows three main findings: (1) CI children and CI adults (with long-term CI use) show lower speech perception accuracy in the presence of background noise in both audiovisual and auditory modalities, but show significantly higher visual speech perception scores, compared to NH peers; (2) CI children’s speech perception during noise is superior to CI adults in the auditory modality, inferior in the visual, and similar in audiovisual integration; (3) Manipulation of context using meaningful and nonsense words did not create difference between CI groups, or between CI and NH participants.

Focusing on each modality separately, auditory speech perception was better for the NH groups than the CIs, and visual perception was better for the CI groups than the NHs. This finding is reflected in the previous literature which found better auditory speech perception abilities in NH individuals as compared to CI users [5, 26], and better visual speech perception among hearing impaired individuals as compared to NHs [23].

Previously, we have shown that the typical development of the visual modality occurs after age 9 [13]. However, in the present study, CI children had better performance in visual perception compared with their NH peers. This may suggest that early deprivation of auditory information affected the CI children’s normative developmental course, and speeded up the development of visual processing, in spite of being implanted at a young age. These findings may indicate that in spite of early implantation and an emphasis on the auditory modality during rehabilitation, CI children use visual information more than their NH counterparts. This outcome is supported by previous studies that have shown hearing impaired children possess better lip reading abilities compared to their normal hearing peers [23, 27-28].
The audiovisual integration in speech perception has been shown in the literature to be better than when using each modality separately, for both CI users and NH individuals [16, 17]. The findings of the current study support these results. However, CI users performed more poorly than NH participants in the audiovisual condition, as well as in the auditory. These results may be explained by the impact of background noise. The presentation of words against noise was chosen in order to avoid a ceiling effect in the NH groups, especially among the NH adults. However, this may have negatively affected the performance of the CI groups to such an extent that visual “compensation” was not enough to equalize their performance with that of their NH peers.

Additional explanation for poor audiovisual performance for CI users might be the use of monosyllabic words. Such words have low visual redundancy that makes recognition of the words (without the complementary auditory information) difficult. Therefore, this task may reduce the typical visual advantage of CI users. Future studies should use more redundant stimuli (such as sentences) in order to clarify this factor.

The findings of the current study also support previous studies showing a developmental course of speech perception in noise, such that NH children have lower speech perception accuracy than NH adults in audiovisual and auditory modalities, in the presence of background noise [13]. However, different pattern was observed for CIs. CI children had higher auditory speech perception than CI adults, and similar audiovisual speech perception. It seems that the CI children better process auditory information, while the CI adults rely more on the visual, in difficult listening conditions. Why do CI children better in auditory speech perception than CI adults when normal hearing individuals show the reverse pattern? One possible explanation
for this reverse pattern relates to the fact that children and adults differ in sensory maturation and linguistic experience. However, when comparing CI children and adults, an additional aspect should be considered: current long term CI adult users were implanted in alternate era than current CI children, an era reflective of different CI candidacy criteria, implantation practices, and technology.

These factors manifest in the present study in a few ways: (1) The CI children were implanted at younger age than the CI adults, which exposes them to auditory information at younger age, which may have a more significant effect than extended experience with speech perception (by virtue of more years of use) that the long-term CI adult users have; (2) All CI children were implanted bilaterally, while half of CI adults were implanted unilaterally; moreover, the adults that were implanted bilaterally received their second implant sequentially, years after the first implant, unlike the CI children who received second implants in closer proximity to the first. An implication of this may be that CI children are better equipped to perceive auditory information than CI adults who may need to rely more heavily on visual information; (3) When the CI adults were implanted, only those with minimal-to-no residual hearing were implanted; today’s implantation criteria also allow children with severe and severe-to-profound hearing loss to receive CI [29]. Therefore, the CI children include those with better baseline hearing ability than CI adults, which provide them hearing advantage over CI adults; (4) CI technology is continuously improving, which enables more recently implanted children the use of more advanced technologies than those previously implanted, which can influence the quality of speech perception. All of these factors may play a role in explaining why CI children perform better in auditory speech perception than CI adults, contrary to NH pattern.
The finding relating to CI adult’s better visual performance than CI children might be explained by both their older age at implantation and their more matured visual system. Since CI adults were initially implanted at older age than CI children, they probably learned to rely more heavily on visual information in their speech perception during their pre-implantation years. Indeed, the literature shows that CI users’ visual brain areas are more developed than their age-matched peers [30]. Their current, older age also provides them with more developed visual system and more experience in using it during speech perception.

1.5 CONCLUSIONS

The results of the current study suggest that in spite of our group of CI children being implanted at a very young age, bilaterally, with better baseline hearing, and using more advanced technology, they still find it difficult to perceive speech in adverse listening conditions (background noise). Adding training in audiovisual perception as a rehabilitation strategy may be beneficial to this group by improving their audiovisual integration, which in turn may help them cope better in difficult listening conditions.

ACKNOWLEDGEMENTS

The authors would like to thank Lital Raifen, Lilach Sabi, Shani Baron, Inbal Brodai, Yardena Goldberg and Avital Natanzon for collecting data for this study.
REFERENCES


http://dx.doi.org/10.1016/j.heares.2014.09.002


http://dx.doi.org/10.1097/00003446-200506000-00003

http://dx.doi.org/10.1121/1.1907309

http://dx.doi.org/10.1044/1092-4388(2010/09-0148)

http://dx.doi.org/10.3766/jaaa.23.6.8


http://dx.doi.org/10.1371/journal.pone.0107252


http://dx.doi.org/10.1006/jecp.1997.2379


http://dx.doi.org/10.1121/1.428656


http://dx.doi.org/10.1111/j.1467-7687.2008.00677.x


http://dx.doi.org/10.1044/2013_JSLHR-H-12-0273


http://dx.doi.org/10.3109/14992027.2014.973536

http://dx.doi.org/10.1348/000709900158272

http://dx.doi.org/10.1097/MAO.0000000000000607

http://dx.doi.org/10.1371/journal.pone.0147793
Table 1. Mean (SD) speech perception accuracy of study groups for meaningful and nonsense words in each modality.

<table>
<thead>
<tr>
<th></th>
<th>CI children</th>
<th>CI adults</th>
<th>NH children</th>
<th>NH adults</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meaningful words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiovisual</td>
<td>45.50 (8.64)</td>
<td>48.50 (20.55)</td>
<td>72.33 (8.84)</td>
<td>86.18 (9.61)</td>
<td>63.13 (12.46)</td>
</tr>
<tr>
<td>Auditory</td>
<td>28.00 (11.83)</td>
<td>18.00 (12.29)</td>
<td>58.33 (14.96)</td>
<td>72.06 (12.25)</td>
<td>44.10 (13.41)</td>
</tr>
<tr>
<td>Visual</td>
<td>9.50 (7.98)</td>
<td>25.00 (8.16)</td>
<td>2.67 (3.20)</td>
<td>5.29 (5.14)</td>
<td>10.62 (6.19)</td>
</tr>
<tr>
<td><strong>Nonsense words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiovisual</td>
<td>36.50 (13.13)</td>
<td>29.50 (10.39)</td>
<td>60.33 (16.63)</td>
<td>75.59 (14.46)</td>
<td>50.48 (14.65)</td>
</tr>
<tr>
<td>Auditory</td>
<td>16.50 (15.64)</td>
<td>3.00 (3.50)</td>
<td>45.00 (16.58)</td>
<td>60.00 (14.03)</td>
<td>31.13 (14.33)</td>
</tr>
<tr>
<td>Visual</td>
<td>5.50 (6.85)</td>
<td>12.50 (8.90)</td>
<td>2.67 (3.72)</td>
<td>5.59 (5.56)</td>
<td>6.56 (6.34)</td>
</tr>
<tr>
<td><strong>Meaningful and nonsense words combined (averaged)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiovisual</td>
<td>41 (9.87)</td>
<td>39 (14.78)</td>
<td>66.33 (9.81)</td>
<td>80.88 (10.19)</td>
<td>56.80 (13.56)</td>
</tr>
<tr>
<td>Auditory</td>
<td>22.25 (13.15)</td>
<td>10.5 (5.24)</td>
<td>51.67 (13.35)</td>
<td>66.03 (10.68)</td>
<td>37.61 (13.87)</td>
</tr>
<tr>
<td>Visual</td>
<td>7.5 (6.12)</td>
<td>18.75 (7.57)</td>
<td>2.67 (2.21)</td>
<td>5.44 (4.26)</td>
<td>8.59 (6.26)</td>
</tr>
</tbody>
</table>

CI- Cochlear Implants, NH- Normal Hearing
Figure 1. Mean speech perception accuracy, averaged for meaningful and nonsense words, of all study groups in each modality

Figure 2. Mean speech perception accuracy of meaningful and nonsense words in each modality, averaged across study groups.