Audiovisual Integration in Cochlear Implant Users

VANDERBILT UNIVERSITY

Iliza M. Butera[1], Ryan A. Stevenson[2], René H. Gifford[1], Mark T. Wallace[1]

Vanderbilt University, Nashville, TN[1], University of Western Ontario, London ON, Canada[2]

COCHLEAR IMPLANT Research Laboratory

Vanderbilt Brain Institute

Introduction

Noisy environments cause decreased auditory speech saliency and an increase in the integration of visual speech cues. Audiovisual integration can lead to a dramatic increase in intelligibility by as much as 15 dB in typical-hearing individuals.

Cochlear implants (CIs) are highly successful neuroprosthetic devices that can enable remarkable proficiency in speech understanding in quiet; however, most CI users still struggle to communicate in noisy environments.

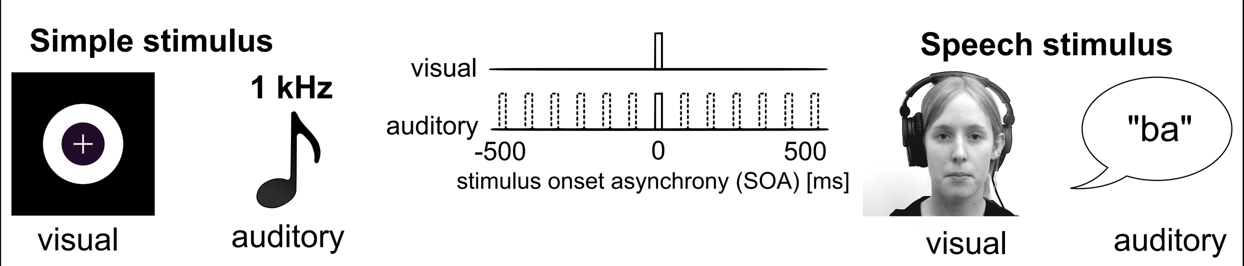
Similar to typical listeners, degraded auditory input provided by a CI is more intelligible when paired with a speaker's visual articulations. The multisensory gain of this interaction is highly dependent on the temporal coincidence of sensory stimuli.

Although much research has investigated the degree to which cochlear implants disrupt the timing, level, and spectral resolution of auditory input, little is known about these parameters in the context of audiovisual integration.

The purpose of this study was to assess temporal parameters of audiovisual integration in adults with cochlear implants and relate this ability to speech proficiency. We hypothesized that the auditory deprivation experienced by CI users would negatively affect their multisensory integrative abilities compared to controls. Furthermore, within the CI cohort, we predicted that lower unisensory and multisensoy thresholds would correlate with higher postoperative auditory-only speech understanding.

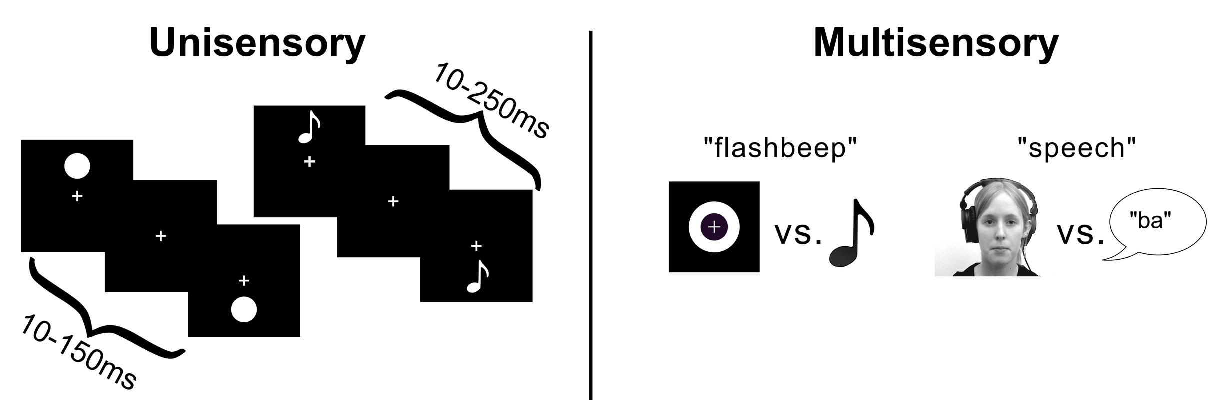
Methods

Simultaneity Judgement (SJ)



Synchronous or Asynchronous?

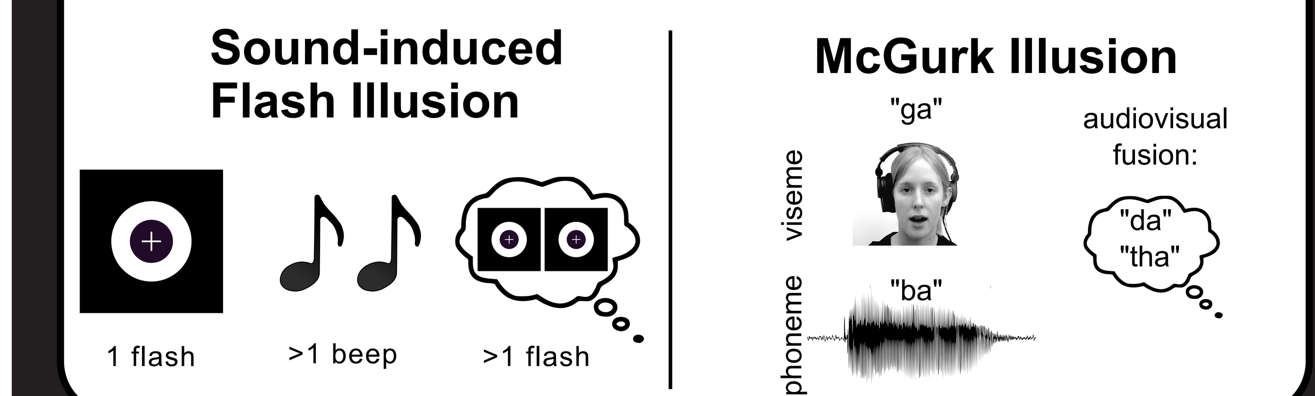
Temporal Order Judgement (TOJ)



Was the higher pitch/circle first?

Was the auditory or visual component first?

Multisensory Illusions



Contact: ilizabutera@gmail.com

Results

Speech scores are representative of the CI population

Group	N	Mean age ± SD (y)	Age range (y)	Number of CIs	Hearing mode	Female (%)
CI	69	59 ± 16	19-86	70% – 1 30% – 2	36% bimodal 64% electric only	51
NH	56	45 ± 17	19-76	N/A	N/A	50

Table 1. 69 Cochlear implant (CI) users and 56 normal-hearing (NH) controls participated in this study. These groups were roughly age-matched; however the influence of age was further investigated in the visual TOJ task with an additional 12 NH controls who were more closely age-matched (mean 55±13y).

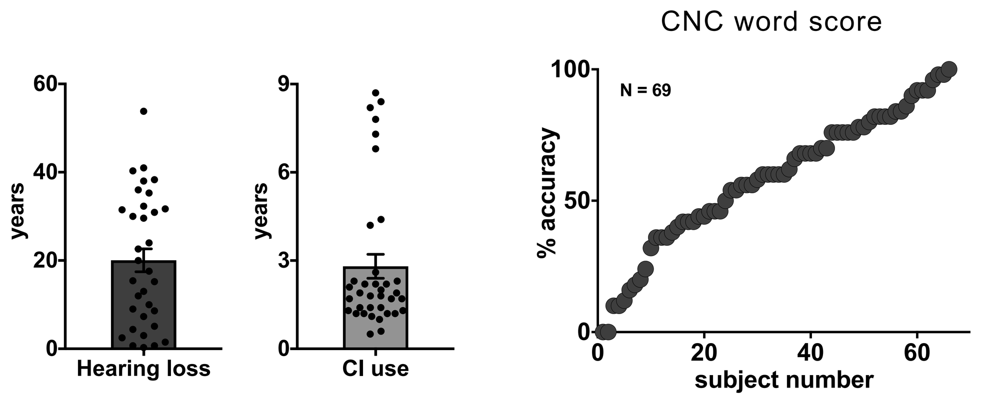


Figure 1. Duration of Hearing loss, CI use, and CNC accuracy

All CI users in this study were post-lingually deafened native English speakers who communicated orally through modern implant processors. Proficiency with speech recognition was assessed via an open set CNC word task as part of the Minimum Speech Test Battery (MSTB). The indvidual data shown above are representative of the broad variability in speech recognition known to plateau following 3-6mo of experience with an implant. Because CI users in this study had an average of three years of experience with their implants, we expect this performance to be maximal for most individuals

CI users' TOJ of multisensory speech is less accurate at large SOAs

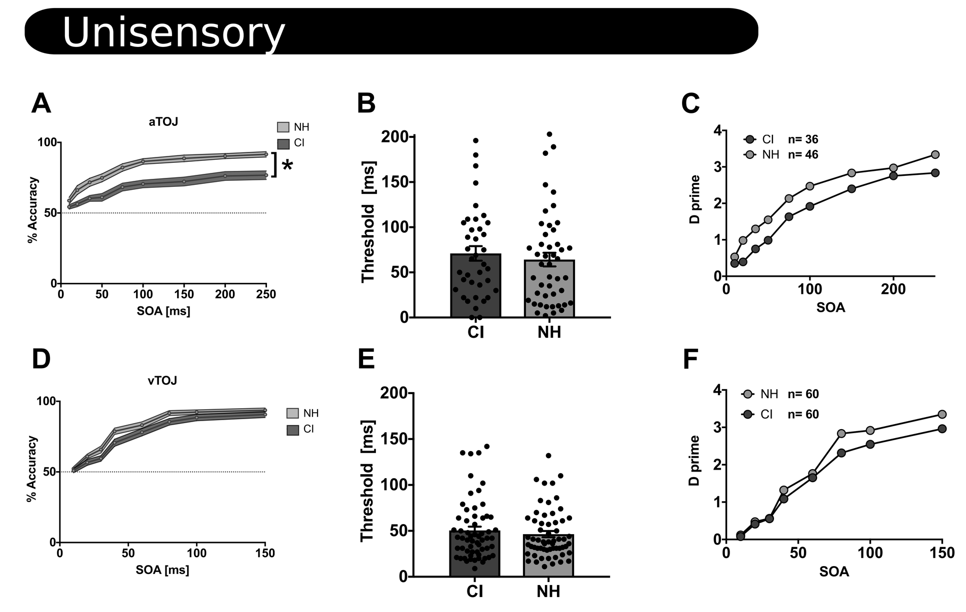
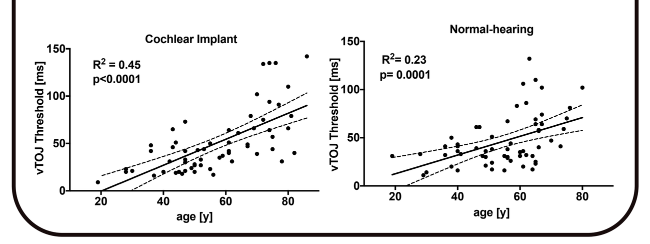


Figure 2. Visual and Auditory TOJ.

(A) Significantly lower performance was seen in auditory TOJ between groups; however after removing subjects' data that could not be fit with standard logit functions (n =23 CI, n=4 NH), thresholds (B) and d prime (C) were not statistically different. Visual TOJ accuracy (D), threshold (E), and d prime (F) were also similar between groups. Shaded area of curves, like all error bars, indicate SEM.*p<0.05

Figure 2 inset.
vTOJ thresholds were the only measure found to correlate with age. After careful age-matching with controls, it appears that older CI users are more susceptable to longer thresholds. After calculating a median split by age and performing a two-way ANOVA, we found no significant effects beyond the strong correlations shown below.



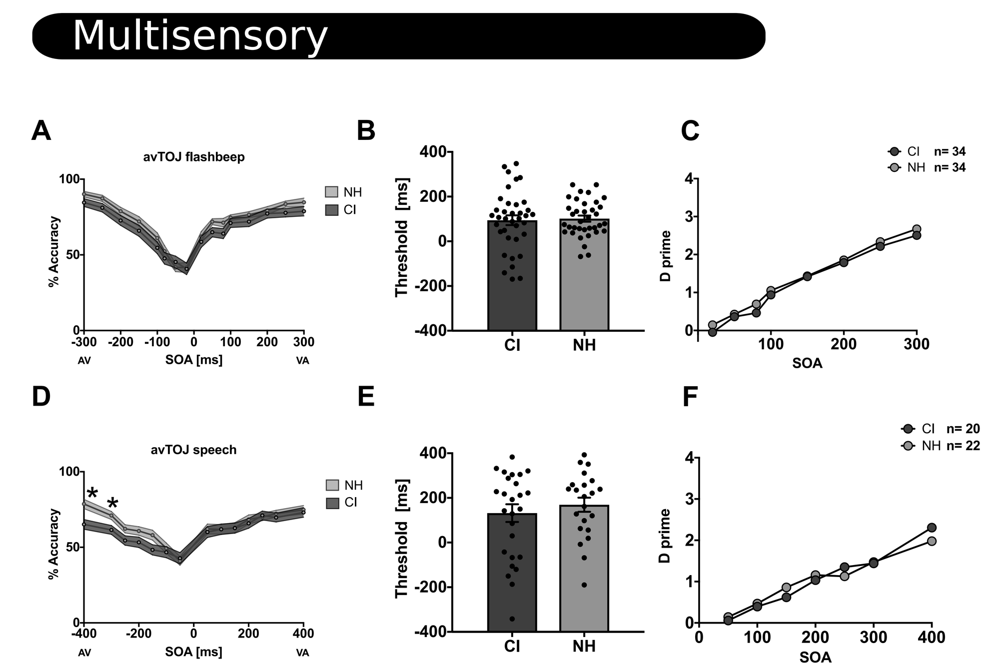


Figure 3. Audiovisual TOJ.

No significant differences in accuracy (A), threshold (B), or d prime (C) were seen in non-speech avTOJ performance between groups. Similarly, speech avTOJ threshold (E) and d prime (F) were comparable; however accuracy was significantly lower at extreme auditory-leading SOAs in the CI group (D). *p<0.05

Perceived synchrony of speech stimuli is greater at large SOAs in CI users

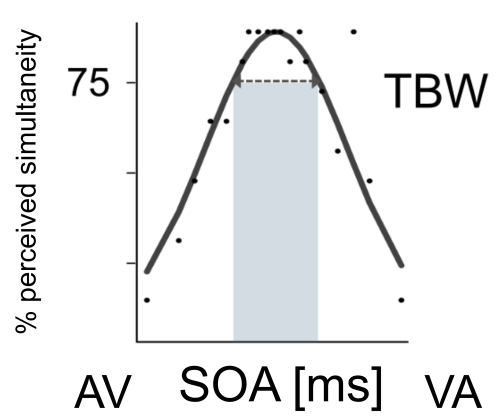


Figure 4. Example Individual Flashbeep Curve
Example psychometric curve from presenting 20 trials per SOA and plotting perceived simultaneity at each asynchrony as well as objective synchrony at 0ms. This was done for both AV stimuli in order to calculate the temporal binding window (TBW) for each task.

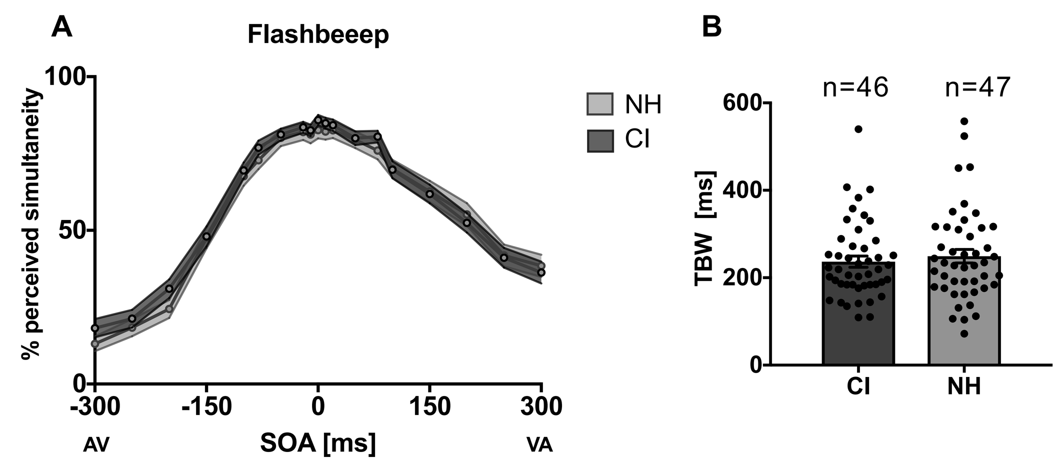


Figure 5. Flashbeep
(A) Average reports of simultaneity at each SOA indicate remarkably similar responses between groups. After omitting subjects' data that could not be fit with standard logit functions (n=8 CI, n=8 NH), TBWs were calculated as the width of psychometric curves at 75% perceived simultaneity. (B) Results indicate no group differences in the mean TBW with these non-speech stimuli.

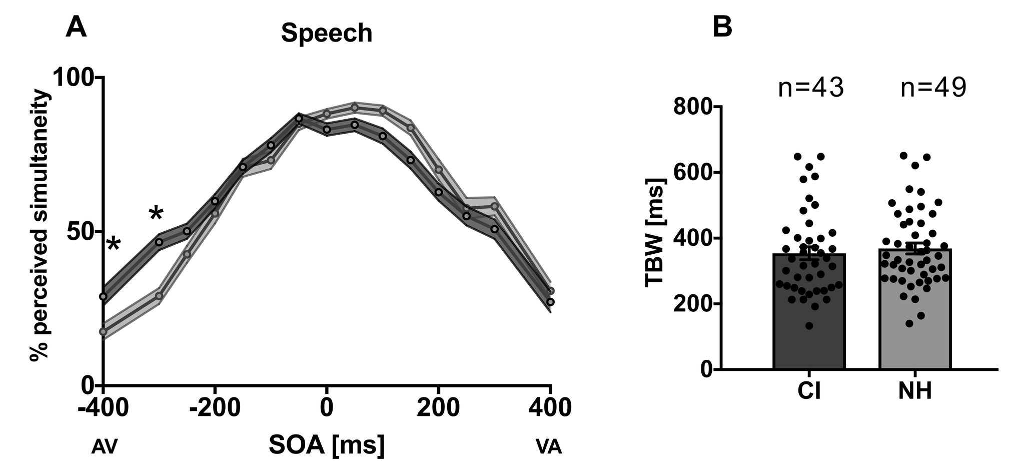


Figure 6. Speech
(A) CI users have significantly greater perceived simultaneity at extreme auditory-leading SOAs. After omitting subjects' data that could not be fit with standard logit functions (n=15 CI, n=4 NH), TBWs (B) were not significantly different. *p<0.05

Results

CI users have lower AV speech accuracy and visual bias with incongruent speech

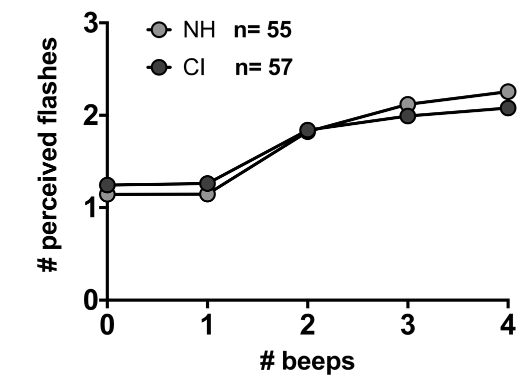


Figure 7. Sound-Induced Flash Illusion

No significant differences were identified between CI and NH groups with either a non-speech illusion or control conditions consisting of congruent numbers of A and V presentations (data not shown).

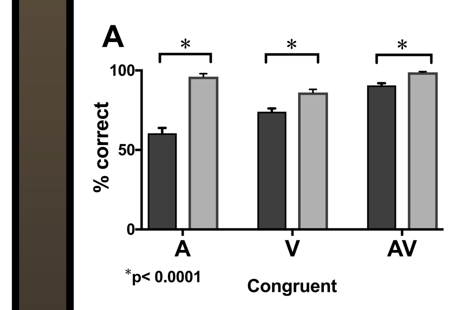
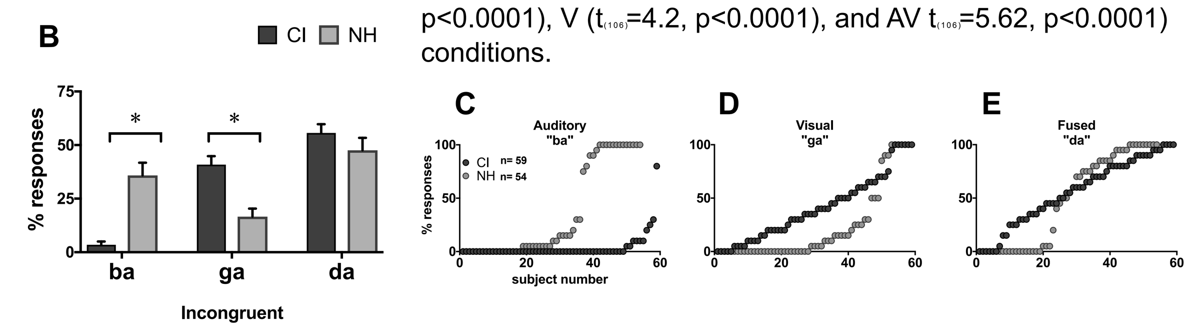


Figure 8. McGurk Illusion

(A) Congruent auditory only, visual only, and audiovisual trials with the syllables 'ba' and 'ga' (combined) indicate lower accuracy for CI users. Repeated measures ANOVA with modality as a within-subject contrast and group as a between-subject contrast indicates that there is a main effect of modality $F_{(2,212)}=31.1$, $p<0.0001$, group $F_{(1,106)}=85.4$, $p<0.0001$, and an interaction $F_{(2,212)}=31.3$, $p<0.0001$. Post hoc pairwise t tests with a Bonferroni correction for multiple comparisons, indicate significantly different means between groups for A ($t_{(106)}=9.0$, $p<0.0001$), V ($t_{(106)}=4.2$, $p<0.0001$), and AV $t_{(106)}=5.62$, $p<0.0001$) conditions.



(B) Incongruent 'McGurk' trials indicate a significant bias toward the visual token ('ga') in CI users ($t_{(111)}=5.47$, $p<0.0001$), a bias toward the auditory token ('ba') in the NH group ($t_{(111)}=4.4$, $p<0.0001$), yet equivalent reports of fusion between the two groups. Individual data for each token is shown in panels C-E.

Summary

In all cases, speech tasks (i.e. TOJ, SJ, and illusions) revealed lower performance in postlingually-deafened CI users that non-speech tasks did not. Interestingly, this was specific to extreme auditory-leading SOAs for both multisensory SJ and TOJ tasks. After excluding subjects for whom psychometric curve fitting was not possible, there were no group differences in auditory-only TOJ thresholds. This finding may indicate difficulties with frequency discrimination in a subset of CI users rather than deficient auditory temporal processing. Additionally, the McGurk task indicated lower accuracy in A, V, and congruent AV perception, as well as visually-biased reports when CI users were presented conflicting A and V information. Because reports of AV fusion were equivalent between groups, it is unclear how effectively the McGurk illusion generalizes to the assessment of AV speech recognition. Finally, none of the tasks in this study were found to correlate with auditory-only CNC word scores. Future work aims to investigate whether temporal thresholds for AV integration generalize to word-level AV speech recognition, particularly in noisy environments where multisensory gain is high even for typical listeners.

Acknowledgment

This study was supported by an NIH T32 MH064913 (Butera), F31DC015956 (Butera), a NSERC Banting Postdoctoral Fellowship (Stevenson), NIH RO1 DC009404 (Gifford), and a Bill Wilkerson Collaborative Grant (Wallace).