Audiovisual listening in cochlear implant users

By
Iliza M. Butera

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Neuroscience

September 30th, 2019
Nashville, Tennessee

Approved:
Troy A. Hackett, Ph.D.
Mark T. Wallace, Ph.D.
René H. Gifford, Ph.D.
Daniel H. Ashmead, Ph.D.
G. Christopher Stecker, Ph.D.

To my parents for many things,
including making me go to school as a kid.
Thanks, you two.

ACKNOWLEDGEMENTS

For the last 6 years (2013–2019), I’ve had the honor of working with Mark Wallace (Neuroscience) and René Gifford (Hearing and Speech), who have served as parallel advisors for my research. Being in both labs has been particularly fun, as I’ve been able to work alongside a diverse group of 40+ students, post docs, and clinicians. Vanderbilt's collaborative environment and plethora of great scientists have enabled me to explore this multidisciplinary research to the fullest extent, and I’m very pleased to share the results here.

I’d like to thank the United States government and US taxpayers, whose shared commitment to scientific progress made this work possible through funding from the National Institutes of Health (T32 MH064913 and 5F31DC015956). And although it usually goes without saying (perhaps from a lack of reciprocation), I’m compelled to share how fortunate I feel to have undertaken this dissertation at a time when many machines—and by extension their firmware, an abundance of software, and of course, the internet—can also support such a scientific endeavor. Special thanks to my personal computer, which has helped me enormously. It’s a modest 13” MacBook Air circa 2012 that, despite its outdated operating system, never lets me down—and for that, I am truly grateful.

Fun facts: this document is a product of 4 years of committee meetings, numerous discussions at 9 academic conferences, nearly 100 1:1 meetings with my advisors, over 200 lab meetings, and ~760 hours of time graciously donated by 202 volunteers. To all my committee members, lab mates, office dwellers, advisors, collaborators, research participants, program administrators, fellow Neuroscience students, family, and friends: thank you all for joining me in this adventure. You’re great!

So many humans have been tremendously supportive, and in the interest of brevity I’ll elaborate on just one: my live-in editor, sounding board, house DJ (as in the place, not the genre), favorite critic, coffee-maker, and begin-and-end-your-day-in-the-best-possible-way kind of person. This dissertation (and likely most things I will do) is a product of the intellectual partnership I’ve developed with Mike Butera. For the last 6 years, Mike’s been a collaborator, mentor, cheerleader (figuratively, not like rah-rah), research participant, friend, and as of my first semester in grad school, family. To all the many (many) others who also wear one or more of those hats, I sincerely thank you all for your general awesomeness and dedication to this fascinating endeavor. Together, you’ve made it so very rewarding to now share its completion. Thank you.

TABLE OF CONTENTS

Chapter

1. An introduction to speech perception with a cochlear implant

2. Visually-biased illusion perception in cochlear implant users

3. Functional localization of audiovisual speech using infrared spectroscopy

4. Behavioral and functional neuroimaging of CI users' audiovisual processing

5. Audiovisual temporal processing in postlingually deafened adults with cochlear implants

6. Discussion and integration of findings

Appendix

LIST OF TABLES

Table Page

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF FIGURES

Figure Page

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF ABBREVIATIONS

CI	Cochlear implant
NH	Normal hearing
ABR	Auditory brainstem recording
FDA	Food and Drug Administration
NIDCD	National Institute on Deafness and Other Communication Disorders
CNC	Consonant nucleus consonant
MSTB	Minimum speech test battery
LNT	Lexical Neighborhood Test
PTA	Pure tone averages
dB SPL	decibels of sound pressure level
db HL	decibels of hearing level
LTASS	Long-term average spectrum
MMSE	Mini-mental state exam
AVT	Auditory-verbal therapy
KBIT	Kaufman Brief Intelligence Test
TONI	Test of Nonverbal Intelligence
ANOVA	Analysis of variance
MANOVA	Multivariate analysis of variance
MANCOVA	Multivariate analysis of covariance
FDR	False discovery rate
SIFI	Sound-induced flash illusion
AV	Audiovisual
p(McGurk)	Probability of McGurk illusion
SI	Susceptibility index
ii	Interactive index
SNR	Signal-to-noise ratio
ECochG	Electrocochleography
EEG	Electroencephalography
fMRI	functional Magnetic Resonance Imaging
PET	Positron emission tomography
CT	Computed tomography
fNIRS	functional Near Infrared Spectroscopy
BA	Brodmann area
MT	Middle temporal
oxyHb	oxygenated hemoglobin
deoxyHb	deoxygenated hemoglobin
ROI	Region of interest
glm	General linear model
STS	Superior Temporal Sulcus
SOA	Stimulus onset asynchrony
SJ	Simultaneity judgment
TOJ	Temporal order judgment
TBW	Temporal binding window
PSS	Point of subjective simultaneity
SDT	Signal detection theory
ROC	Receiver operating curve
NED	Noisy encoding of disparity

CHAPTER 1 An introduction to speech perception with a cochlear implant

1.1 Cochlear implant function and candidacy

The cochlear implant (CI) is the single-most successful neuroprosthetic device available today, restoring auditory function to those who are profoundly deaf. Unlike hearing aids, which simply amplify sounds, CIs circumvent the middle ear entirely by processing sounds externally. More specifically, the device consists of two halves: an external processor and an internal implant, held together by a transdermal magnet. Sounds are detected by a microphone near the opening of the ear canal and the processor worn behind the ear translates the frequency and amplitude of sound waves into electrical pulses. This digital signal is then sent to an internal receiver via a radio frequency coil. Lastly, the electrical pulses are delivered to the inner ear, where surviving auditory spiral ganglion cells in the cochlea are excited in a tonotopic manner (i.e., from high to low frequencies, spiraling inward). This entire process takes place in approximately 10ms and establishes a rudimentary sense of hearing to those with a severe impairment.

In 2012, the National Institute on Deafness and Other Communication Disorders (NIDCD) estimated the number of worldwide CI users to be over 324,000 (Fig. 1.1a). Given the more recent addition of emerging markets, such as in China and India, actual quantities today are likely to exceed half a million individuals. Indeed, based on device registrations with implant manufactures, today there are over 100,000 CI users in the US alone, many of whom are pediatric recipients (Fig. 1.1b).

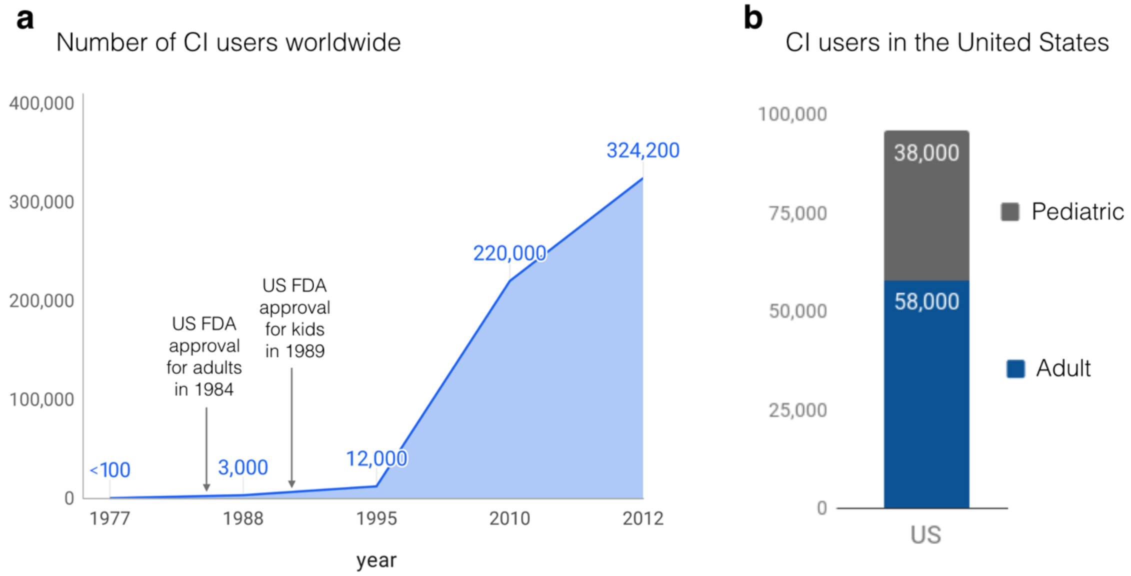


FIGURE 1.1. Worldwide cochlear implant use over time. (a) Beginning with the seminal clinical trials of the late 1970s, the number of CI users has grown to over 300,000 people worldwide. (b) In the United States alone, there are currently around 100,000 CI users.

For adults, CI candidacy is typically determined by pure-tone hearing thresholds designated as severe-to-profound hearing loss in the range of 70 to 90 dB HL. When initial approval by the United States Food and Drug Administration (FDA) was granted, there were strict guidelines that required essentially no acoustic hearing whatsoever, regardless of frequency. Today the criteria are much more lenient, and it is not uncommon for users to have some residual hearing, particularly in low frequencies (Gifford et al., 2017). Furthermore, many CI users now opt to implant both ears. In 2008, for example, there were approximately 8,000 bilateral CI users, which made up 5% of the worldwide population at the time (Peters et al., 2010). Thus, without profound bilateral sensorineural hearing loss as a prerequisite, the adult CI population now contains substantial clinical diversity in regard to etiology, amount of residual hearing, and the number of implants per person.

For prelingually deafened patients such as infants, it can be challenging to quantify hearing loss objectively (i.e., without the patient's communication of hearing thresholds in an audiogram). Fortunately, in normal development, the inner ear is fully-functional by the third trimester of pregnancy, and auditory brainstem recordings (ABRs) can quantify subcortical auditory-evoked activity at birth. With comparisons to standard neonatal waveforms, mandatory newborn hearing screenings can diagnose congenital hearing loss in the very first postnatal days. While there are clear benefits of early hearing intervention for congenital deafness, cochlear implantation is an irreversible surgery and premature intervention can risk more extensive inner ear damage from the electrode insertion itself. After much research and oversight of the risks and benefits, U.S. FDA approval for pediatric implantation at 12 months of age has been in place since the year 2000. As a result, early pediatric CI surgeries are frequently performed and are associated with beneficial outcomes later on. These benefits include typical development of cortical auditory evoked potentials and meeting broader language milestones throughout childhood (Marschark and Hauser, 2008; Marschark et al., 2007; Wang et al., 2008), provided that implantation is sufficiently early (Gilley et al., 2008).

1.2 Clinical outcome measures and the time course of hearing restoration

Cochlear implants are designed to restore functional aural communication to recipients, so the primary outcome measures for postoperative success are quantified via speech testing. As implant technology has improved over time, new measures have been developed in order to accommodate higher patient performance (Gifford et al., 2008). The standardized Minimum Speech Test Battery

(MSTB) now includes: open-set word recognition (CNC; Peterson and Lehiste, 1962), sentence recognition (AzBio; Spahr et al., 2012), and speech-in-noise testing (BKB-SIN; Killion et al., 2001) to characterize aural listening proficiency.

A period of acclimation is expected for both an audiologist to refine an individual's personalized programming and for the recipient to adapt to the unique sound that an implant conveys. The typical time course of this process begins with initial device activation, typically one month following surgery. The aforementioned clinical assessments track the hearing restoration that takes place over the next several months, with asymptotic performance typically after 3-6 months and at a broad range of final possible outcomes (Fig. 1.2).

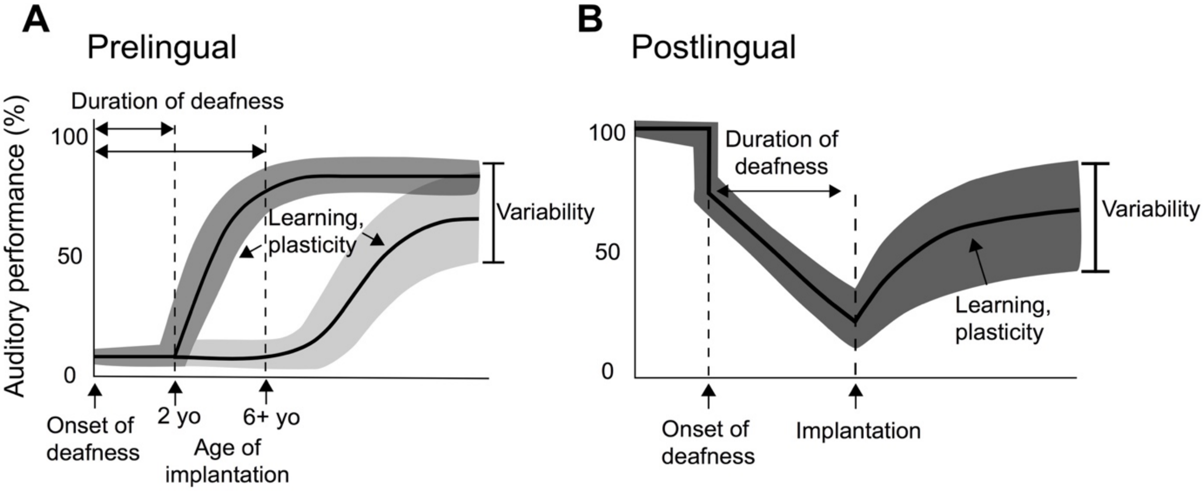


FIGURE 1.2. Schematic of auditory (re)habilitation with cochlear implants. Following (a) prelingual or (b) postlingual onsets of profound hearing loss, auditory performance typically plateaus between 3-6 months post implantation with considerable individual variability represented by shaded areas. Adapted in part from (Blamey et al., 2013; Holden et al., 2013; Lazard et al., 2012a)

Despite the widely-acknowledged success of these devices, the hearing that they enable is quite unlike what is experienced by acoustic listeners. Furthermore, the low spectrotemporal resolution of cochlear implant-mediated sound presents a substantial limitation for deciphering complex sounds like music and speech-in-noise. As a result, even top performers on speech testing in quiet environments can have much lower success in ecological listening situations.

1.3 Spectrotemporal resolution of cochlear implants

Early prototypes of cochlear implants contained only one informational "channel" to convey what was essentially temporally-correlated broadband noise. Despite the inability to discern any meaningful speech with these devices alone, early test subjects had improved detection of environmental sounds and actually even performed slightly better while lipreading (Bilger and Hopkinson, 1977). In this and other early studies, participants also reported less effortful communication and, generally, a preference for even this very rudimentary form of auditory stimulation rather than none at all (Wilson and Dorman, 2012).

Over the last forty years, many technological and surgical advancements have occurred in the field of cochlear implants. For instance, there are now between 8–22 channels in a standard electrode array in addition to improvements in signal processing. Hybrid hearing aid technology can also allow users to combine electric and acoustic hearing in one ear. And on the surgical side, improved insertion methods and the development of flexible electrodes has also minimized intraoperative trauma to the inner ear. Despite these significant advancements, there are still physiological limitations preventing cochlear implants from conveying sounds at the high spectrotemporal resolution of acoustic hearing.

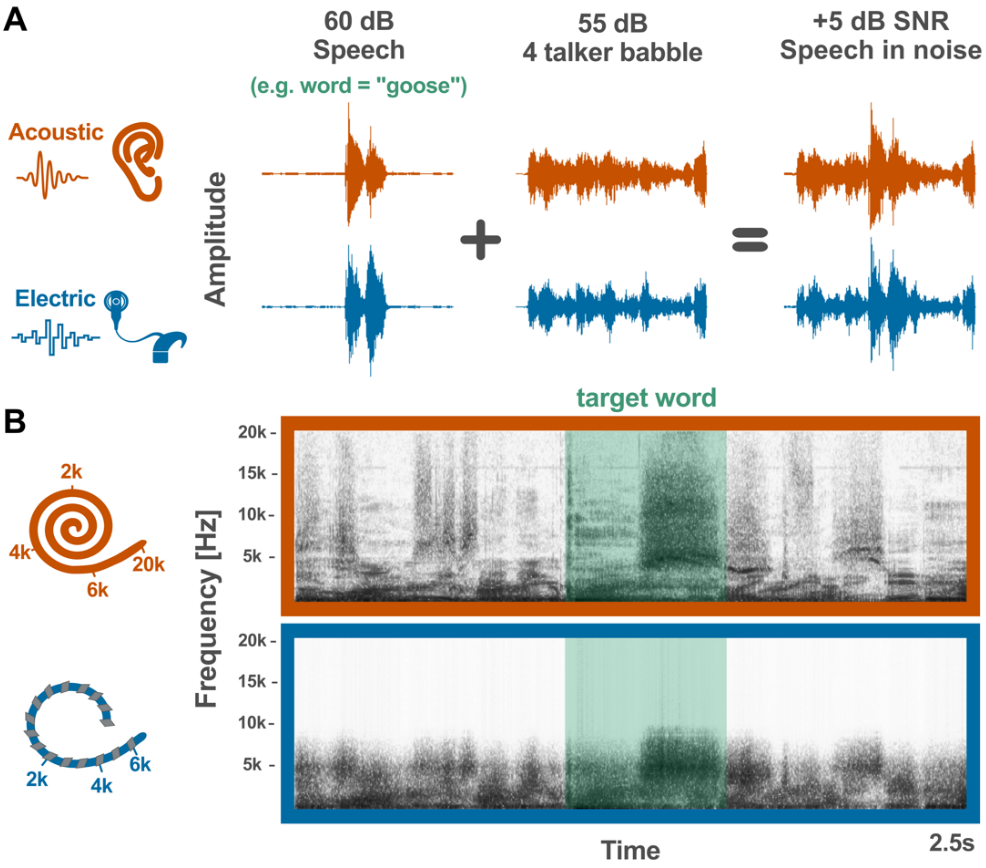


FIGURE 1.3. The low spectrotemporal resolution of cochlear implants makes it challenging to listen in noise. (a) While the temporal components of the acoustic waveform of a word is well-conveyed by an implant, (b) the spectral information is compressed and more difficult to perceive.

The electrode-neural interface in the inner ear presents several challenges. Most notably, because the inner ear is a fluid-filled cavity, the electrical pulses at each contact point on the electrode array stimulate the immediately-adjacent spiral ganglion cells (as intended) but also conduct through the surrounding perilymph to more distant cells. This spread of activity causes channel interaction such that using 12 electrodes, for instance, does not actually represent 12 discrete channels of data. Functional channels are probably closer to 7, regardless of the number of electrodes on the array (Friesen et al., 2001).

To illustrate the consequences of channel interaction, the overall waveform of a word plus a mild signal-to-noise ratio (SNR) is plotted in orange (Fig. 1.3a) along with the same sounds filtered through a cochlear implant processing algorithm in blue [developed at Advanced Bionics by (Litvak et al., 2007)]. Because CIs excel at conveying temporal cues, the overall waveform of acoustic speech (orange) is well conveyed in the electric signal, plotted here via an implant processing algorithm (blue). In a +5 dB SNR—a common and relatively easy signal-to-noise ratio for normal-hearing listeners—there is minimal energetic masking of a monosyllabic target word by competing background talkers. However, plotting this "speech-in-noise" example in a spectrogram (Fig. 1.3b) illustrates the restricted frequency range of a CI (blue), which is conveyed at a lower spectral resolution than acoustic speech (orange). As a result, identifying speech-in-noise is far more difficult for CI users, who consequently rely more on visual cues to compensate.

1.4 Known predictors for variable outcome measures

Though the CI is the most successful intervention for sensorineural hearing loss, the degree of benefit it confers is highly variable and difficult to predict at the individual level. Postoperatively, most CI users are able to achieve satisfactory speech recognition scores in quiet environments [e.g., (Gifford et al., 2014; Holden et al., 2013; Runge et al., 2016)]; however, most still report high listening difficulty (Gifford et al., 2017). Furthermore, performance may range from 0 to 100% for both postlingually and prelingually deafened adults [e.g., (Blamey et al., 2013; Duchesne et al., 2017)], and further explaining this enormous variability remains an area of intense inquiry (Giraud and Lee, 2007; Fitzpatrick et al., 2014).

These efforts can be subdivided into the following four areas:

(1) Clinical predictors—At present, duration of deafness is the most predictive clinical (i.e., non-surgical) measure, which may account for up to 22% of the postoperative variability in speech outcomes (Lazard et al., 2012b). Although several studies have shown significant correlations

between duration of deafness and postoperative speech recognition performance (Blamey et al., 1996, 2013; Friedland et al., 2003; Rubinstein et al., 1999), it is worth noting that these correlations are largely driven by the most extreme ends of the function. That is, the shortest durations of deafness (< 1 year) are associated with the best postoperative speech understanding and the longest durations of deafness (> 20 to 30 years) are associated with the poorest postoperative speech understanding (Blamey et al., 1996, 2013; Friedland et al., 2003; Rubinstein et al., 1999). For kids, age of implantation is typically the best predictor of performance with earlier implantation being better (Ching et al., 2014; Niparko et al., 2010). Residual hearing is also a positive predictor for higher success in adults and children who can augment the rudimentary electrical signals of an implant with some acoustic input.

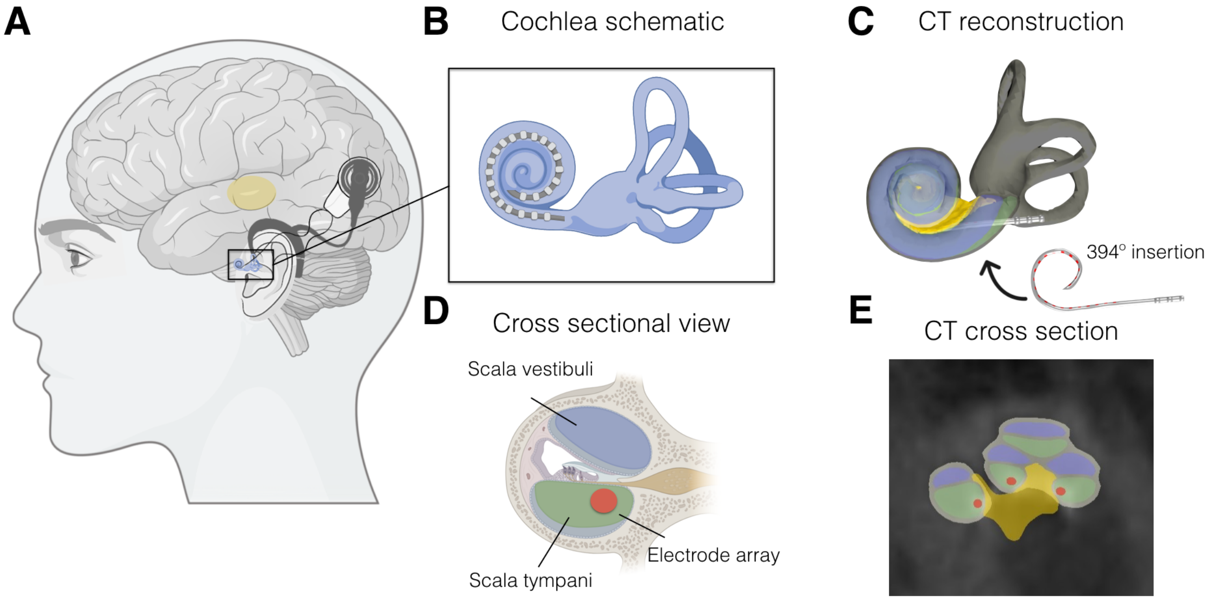


FIGURE 1.4. An overview of optimal electrode placement in the cochlea. (a) The external processor (gray) transmits electrical pulses to an internal receiver and an electrode array within the cochlea (blue), which causes auditory spiral ganglion cells to fire action potentials and ultimately transmit information to the auditory cortex (yellow). (b) Depending on the manufacturer, the electrode array in the cochlea can have up to 22 channels. (c) For a patient with a Cochlear™ Nucleus® CI532 electrode array, a CT reconstruction indicates full insertion just over 360°. (d) A cross sectional view of the cochlea indicates the ideal placement of the electrode array within the scala tympani. (e) For this patient, the electrode array (red) is indeed within scala tympani (green) throughout multiple turns in the cochlea as depicted in this labeled section of a high resolution temporal bone CT scan. Figure panels c and e were created using the methods described in (Noble et al., 2013); CT= computed tomography

(2) Surgical predictors—Electrode scalar location, electrode-to-modiolus distances, and angular insertion depth have all been demonstrated to impact CI outcomes (Labadie et al., 2016; Noble et al., 2014, 2016). Patients have the highest outcomes when electrodes are inserted in scala tympani (without any crossover to scala vestibuli; Fig. 1.4) and without tip foldover or over/under insertion (Finley et al., 2008; Wanna et al., 2014; Zuniga et al., 2017). All these factors contribute to the degree of channel interaction, because the further electrode contacts are from their neural targets, the greater the amount of current is applied and, consequently, spread throughout the cochlea.

(3) Peripheral pathophysiology—Spiral ganglion survival may be inconsistently distributed within the cochlea, causing "dead regions" for some frequencies. Because CIs rely on the integrity of these neural structures, peripheral nerve atrophy is a significant predictor of variable speech comprehension (Fitzpatrick et al., 2014). A functional measure of spiral ganglion survival can be assayed intraoperatively with electrocochleography (ECochG). Doing so can identify these dead regions where no current should be directed upon implant activation. Additionally, ECochG can be used for real-time feedback in the operating room to indicate when a full insertion has been reached (Koka et al., 2018; Riggs et al., 2019). Postoperatively, a CI-mediated sound may be delivered to as few as 15% of the typical number of spiral ganglia (Nadol et al., 2001). Furthermore, subsequent auditory processing is then carried out by either naïve or remodeled primary and associative cortical areas as a result of prior sensory deprivation.

(4) Central Pathophysiology— Sensory deprivation is likely to impact both distal and proximal neural populations in ways that may independently or sequentially inhibit the interpretation of the noisy auditory information that a CI delivers. *An important question for addressing the variability in implant success is identifying where in the auditory pathway variability is introduced.* This dissertation addresses the role of brain plasticity in the central nervous system with a focus on multimodal sensory integration.

1.5 Clinical relevance of multisensory processing for CI users

Listening in the presence of background noise—a common, perhaps daily, experience—is exceedingly difficult for most cochlear implant users. In fact, the degraded spectral content that cochlear implants convey can make listening, even in a quiet room, both challenging and effortful for many users. Fortunately, visual input can improve auditory thresholds (Barone and Deguine, 2011; Barone et al., 2010; Grant and Seitz, 2000b). Audiovisual integration is the process of filtering and relating sensory inputs, like auditory speech and visual articulations, in order to create a more accurate percept than either sense could on its own (e.g., just lipreading or talking over the phone). Each sensory modality encodes complementary information that is increasingly useful when their saliency/interpretability is relatively low. This principle is called "inverse effectiveness." While typical, acoustic listeners may only experience the benefits of audiovisual integration in, for example, a very noisy restaurant, electrical hearing through a CI may in itself introduce enough "noise" to necessitate the integration of visual cues as a constant, compensatory strategy. In effect, audiovisual integration may be required for CI users to successfully understand speech and converse in any listening environment, not just a noisy one.

The synergistic use of auditory and visual information is well-established to confer significant advantages. These include increased saliency (Stein et al., 1996), decreased detection thresholds (Grant and Seitz, 2000a; Lovelace et al., 2003a), and reduced reaction times (Gilley et al., 2010). Naturalistic oral communication is typically an audiovisual (AV) experience that reveals striking perceptual benefits in speech comprehension when visual cues are present (Sumby and Pollack, 1954). For CI recipients, the mechanisms of multisensory integration are likely to be unique in that they occur following a period of sensory deprivation that may have altered brain organization. Indeed, the absence of auditory input during critical developmental periods has been suggested to result in a failure of multisensory integration to develop [e.g. (Schorr et al., 2005)], with consequent effects on speech understanding (Lachs et al., 2001).

1.6 Intramodal and crossmodal plasticity

There has been a considerable amount of work investigating the impact of deafness and blindness on spared modalities (Merabet and Pascual-Leone, 2010). For instance, in the absence of any actual visual stimulus, reading braille elicits activity in the visual cortex of the blind (Reich et al., 2011). In a similar way, the left auditory cortex of the deaf is active during sign language—a phenomenon not present in hearing signers (Calvert et al., 1997; Lee et al., 2007). These are examples of so-called "crossmodal" plasticity, because the cortical areas typically dedicated to visual and auditory processing are underutilized in sensory-deprived brains and, as a result, are co-opted by other modalities—in this case, tactile and visual inputs.

In addition to *cross*modal changes, plasticity may be defined as *intra*modal or occurring simply within the spared modalities. That is, the deaf visual cortex in itself may be remodeled to some extent. For instance, there are several visual tasks associated with enhanced behavioral performance during deafness. These include movement detection thresholds, detection reaction times, peripheral motion processing, visual-only speech understanding (i.e., lipreading), and peripheral attentional modulation (Bottari et al., 2010; Chen et al., 2006; Dye et al., 2009; Lore and Song, 1991; Neville and Lawson, 1987a, 1987b; Reynolds, 1993). Mechanistically, these adaptive changes may occur either within visual, auditory, or multisensory cortical areas to functionally compensate for sensory deprivation.

1.7 Functional and structural correlates of brain plasticity during deafness

Several functional correlates of deafness-related brain plasticity have been identified for both simple tasks and language-associated tasks. In one of the first examples, visual motion was shown to recruit right auditory cortex more in the deaf than in hearing controls (Finney et al., 2001). Extending from this, exposure to sign language alone was found sufficient to alter the lateralization of area MT, an area specialized for visual motion processing, from the right hemisphere to the left (Bavelier et al., 2001; Bosworth and Dobkins, 1999; Fine et al., 2005). In addition, increased attention to visual motion recruits posterior STS (pSTS) in deaf signers but not hearing controls or hearing signers, illustrating that early sensory deprivation (and not signing) may alter inputs to brain regions responsible for audiovisual (multisensory) integration. In addition to hemispheric differences, greater visually-evoked potentials in the deaf correlate with faster RTs to detect peripheral stimuli (Neville and Lawson, 1987b). Furthermore, deaf signers exhibit right visual field advantages for motion as well as form processing, suggesting that dorsal and ventral pathways are left hemisphere dominant in deafness (Bosworth and Dobkins, 2002).

Activation patterns during active sign gesturing in deaf individuals are also similar to hearing controls performing a similar task in spoken English (Petitto et al., 2000). This is further evidence that auditory areas remain task-dedicated (e.g., to language processing) despite the absence of auditory input such that vision alone can elicit similar activation patterns. Thus, compensatory language processing through intact sensory systems may become increasingly lateralized toward language processing networks regardless of the typical sensory input.

Reports of enhanced peripheral attention during deafness suggest that connectivity between motion-selective visual areas and attention networks of the parietal cortex may be altered as well. To this point, Bavelier and colleagues (2000) demonstrate effective connectivity between MT/MST and the posterior parietal cortex to be stronger in deaf than hearing controls in conditions of peripheral attention. In one of very few papers that indicate a structure-behavior correlation regarding crossmodal plasticity in deafness, Shiell and colleagues (2016) report that greater cortical thickness in the right planum temporale in deaf individuals is positively correlated with better visual motion detection thresholds. From this body of work, we expect structural and functional differences in auditory, visual, and audiovisual brain regions to relate to post-implantation audiovisual speech performance.

1.8 Compensatory versus maladaptive views of the role of vision

Research of developmental sensory plasticity has a long history. Since the seminal work of Hubel and Wiesel of the 1960s and 70s, we now know quite a bit about critical periods in development. In these early and sometimes short epochs, sensory input must be experienced in order for normal neural architecture and function to develop. Cochlear implants are the first piece of technology able to interface with a human sensory system and restore a deafferented modality. As a result, this sensory augmentation presents a new application of previous basic research of sensory development to a clinical population. That is, new clinically-focused questions are emerging regarding the impact of deprivation-induced changes on later hearing rehabilitation through CIs (Stropahl et al., 2017a).

It should be noted that although retinal implants are an analogous technology to CIs, there are several unique technological barriers, and their implementation is far behind hearing technology. Instead, a more relevant comparison of visual restoration is cataract removal surgery. Particularly in developing countries, access to medical care may be limited and result in severe and long-term visual impairment despite effective surgical solutions for this condition. In extreme cases of highly-obscured vision during critical periods, cataract removal can be disadvantageous for patients who lack the experience-driven synaptogenesis that strengthens and refines visual processing pathways. Consequently, some of these formerly visually-impaired individuals struggle to functionally make use of restored vision and are better adept with continuing to relying on auditory and tactile inputs.

An overly-simplistic dichotomy has emerged in the literature regarding deafness-related plasticity as being either compensatory or maladaptive. These contradictory views stem from studies that relate crossmodal plasticity to behavioral advantages [reviewed in (Voss et al., 2010)], while others suggest that poor outcomes are linked to the persistence of crossmodal activations that do not revert to the typical functions in reafferented sensory cortex (Collignon et al., 2011; Kral and

Sharma, 2012; Sandmann et al., 2012; Sharma et al., 2009; Voss, 2013). In reality, the impact of deafness-related plasticity on hearing restoration is a complex and multifaceted issue that is unlikely to fit into an adaptive/maladaptive dichotomy (Heimler et al., 2014).

Clinically, however, there are therapies that essentially put into practice the view of vision as maladaptive for hearing restoration. These therapies discourage CI users from lipreading, and instead emphasize strict auditory-based rehabilitation. This approach called Auditory-verbal therapy (AVT) essentially treats visual speech cues as a sensory crutch that needs to be removed in order for a therapist to strengthen a patient's nascent sense of hearing and not merely their lipreading proficiency. This strategy may be underestimating the potential for combined audio-visual rehabilitation, while broadening this intervention to the visual domain might extend perceptual benefits. Recent work in animal models has begun to provide more support for the benefits of vision and its more likely role as a compensatory mechanism for CI users.

1.9 Neural mechanisms of crossmodal plasticity in animal models

Mechanistically, it is feasible that direct inputs from the peripheral visual field of V1 to A1 underlie dorsal visual motion enhancements identified during deafness. These direct projections between A1 and peripheral visual fields of V1 were first uncovered using tract tracing in a non-human primate model (Falchier et al., 2002). Later work in the deafened cat showed physiological and functional remapping of auditory regions to process visual information in visuospatial orienting (Meredith et al., 2016). Importantly, these effects appear to have a high degree of regional specificity. For example, peripheral visual localization in deaf cats involves the same regions that hearing cats require for accurate auditory localization. Furthermore, reinstatement of hearing through a cochlear implant in these animal models causes this cross-modal reorganization to immediately revert to a pre-deafening natural state (Land et al., 2016). Also, at the neuronal level, animal studies have shown that deafening results in dramatic increases in the number of multisensory neurons in auditory areas, with these neurons showing visual and/or somatosensory responsivity (Meredith and Lomber, 2011). The functional role of these multisensory neurons in subserving visual and audiovisual behavior in humans has yet to be determined.

As a result, one outstanding question in the literature is: do the plastic changes that occur with a deafferented sensory cortex in humans inhibit later sensory restoration? That is, if auditory cortex is repurposed for visual functions, can it revert back to typical auditory processing via a cochlear implant, or are these neural changes detrimental to sensory rehabilitation?

1.10 Clinical relevance of crossmodal plasticity for CI recipients

Several studies have provided insight on neural profiles of high and low performing CI users. For example, increasing intelligibility of speech stimuli has been correlated with larger auditory cortical responses in normal-hearing controls as well as in high-performing CI users (Olds et al., 2016). In contrast, those with low speech perception scores exhibit broad activation profiles without preference for intelligible speech. An fMRI investigation of CI candidates as they evoke auditory imagery of either speech or environmental sounds, showed better post-operative outcomes with typical right lateralization of environmental sound imagery (Lazard et al., 2013). Furthermore, duration of deafness corresponded to lower activity in the left posterior superior temporal sulcus (STS) to phonological speech cues and in the right pSTS to environmental sounds. From this study, functional reorganization seems to increase with longer durations of deafness, leading to an overall reduction in temporal lobe activation, which may affect later CI success.

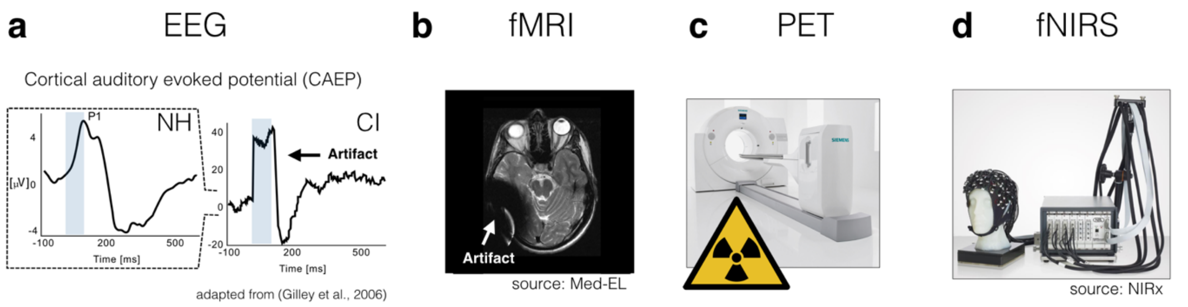
In a related manner, several studies suggest that decreases in auditory cortical activation may be inversely related to visual "takeover" of auditory cortex. For instance, hypometabolism of auditory areas prior to implant surgery has been linked to better post-operative speech outcomes (Lee et al., 2001). Additional PET and EEG studies further reinforce that visual recruitment of auditory cortex, particularly in the right hemisphere, has a detrimental effect on speech outcomes following CI (Doucet et al., 2006; Sandmann et al., 2012; Strelnikov et al., 2015a). Based on this collective work, we suggest that the establishment and strengthening of left hemisphere speech networks through vision during deafness may be beneficial, whereas plasticity more lateralized to the right hemisphere may offer no, or actually impede, these benefits.

Further work in postlingually deafened CI users has indicated that high functional activity in visual cortex is positively related to speech outcomes (Strelnikov et al., 2015b, 2015a). These findings are more in line with the notion that visual proficiency could serve a compensatory role, particularly in multisensory speech recovery. A PET study of speech processing reported that CI users have a larger contribution of visual cortex to speech recognition that's positively correlated with lipreading ability (Giraud et al., 2001). Higher auditory cortex activation during a visual discrimination task is also positively related to face recognition abilities in postlingual CI users, and it is feasible that the high ecological relevance of faces for speech could selectively guide beneficial plasticity (Stropahl et al., 2015). As the severity of hearing loss increases, it seems that functional connectivity between auditory cortex and right MT also increases (Puschmann and Thiel, 2017). Together, these findings lead us to hypothesize that functional imaging will reveal the activation patterns and networks that correlate with visual behavioral enhancements, assessments of AV gain, and auditory proficiency more broadly.

It is also noteworthy that functional and structural measures of crossmodal plasticity, particularly in older work, may have been influenced by long durations of deafness and sign language fluency that is no longer representative in the CI population. Thus, prior studies describing “maladapative” effects of crossmodal plasticity may only represent very low performing CI users for whom we already know duration of deafness is a significant predictor of outcomes. It is likely that intermediate durations of deafness, particularly with sufficient oral language experience, may produce less extensive crossmodal recruitment of right hemisphere auditory areas and possibly improved connectivity between auditory and visual areas in support of multisensory integration.

1.11 Challenges for neuroimaging in cochlear implant users

In order to answer these mechanistic questions of the benefits or consequents of deafness-related brain plasticity, there are several practical considerations for how to do neuroimaging in CI users. Encephalography (EEG) is a common non-invasive imaging technique that detects very small electrical signals generated from neural activity and recorded at a high temporal resolution. Unfortunately, the electrical components of the CI processor and the signal it transmits create massive artifacts (Gilley et al., 2006; Wagner et al., 2018). Several labs are working to develop better filtering methods to remove these artifacts from cortical auditory-evoked potentials (Fig. 1.5a), though success in doings so is limited to very short stimuli like synthetic phonemes. Until further advances, it is not possible to study speech recognition at a natural timescale of even short, monosyllabic words via EEG.



	EEG	fMRI	PET	fNIRS
PROS:	High temporal resolution	High spatial resolution	Compatible with CIs	Compatible with CIs
CONS:	CI electrical artifacts	CI safety concerns	Uses ionizing radiation	Limited spatial resolution of only the top 5-8 mm of cortex

Fig. 1.5. A comparison of neuroimaging techniques. (a) Electrical artifacts in EEG from cochlear implants occur immediately after a stimulus is presented (blue rectangles). These artifacts are very large (~10x the voltage amplitude of the CAEP) and difficult to filter out. (b) fMRI interacts with the ferromagnetic components of CIs and presents safety concerns. (c) PET is compatible, yet is more invasive on account of added contrast agents. (d) fNIRS is a suitable non-invasive technique with the primary downside of limited spatial resolution. (Gilley et al., 2006)

Post-operative fMRI is also not possible on account of the ferromagnetic components of implants. These components introduce potential safety concerns of shifting electrodes, demagnetizing implants, and excessively heating surrounding tissues (Majdani et al., 2008). Furthermore, the implanted components themselves create such large artifacts in the magnetic field that nearly an entire hemisphere can be obstructed by one implant (Fig. 1.5b). Although some MRI-compatible implants are now available, the aforementioned safety concerns do not outweigh the benefits for research purposes alone.

As a somewhat hybrid approach between EEG and fMRI, magnetoencephalography (MEG) is possible to use with CI users; however, it requires a magnet-free prosthesis and a radio frequency shield (Pantev et al., 2006). In research more generally, MEG is less common on account of the very complex and costly building requirements for appropriate shielding, with most labs opting instead for fMRI. Positron emission tomography (PET), however, does not require such shielding and has been used in several CI studies of neural processing (Strelnikov et al., 2015a). The downside to this technique is its invasiveness given that radioactive isotopes must be injected intravenously.

Much like fMRI, functional near infrared spectroscopy (fNIRS) is used for noninvasive, functional imaging of hemodynamic responses in the brain (Fig. 1.5d). Emitters are tuned to wavelengths in the near infrared light spectrum in order to measure adsorption resulting from fluctuating concentrations of oxygenated and deoxygenated hemoglobin in tissues. This serves as a bloodoxygen-level dependent (BOLD) signal from which physiological processes can be inferred. Unlike the indirect measure of disturbances in a magnetic field by these species in MR-derived BOLD signals, NIRS is a direct measure of their concentrations at a much higher sampling rate (usually around 10 Hz, depending on number and arrangement of emitters). The use of fNIRS also has several unique advantages over other imaging techniques including: silent operation, greater resistance to movement artifacts, insensitivity to electrical interference, compatibility with implants, and oftentimes, portability. NIRS is safe to use in many clinical populations including neonates, others with implanted medical devices such as deep brain stimulation devices, as well as people who may be uncomfortable or otherwise ineligible for neuroimaging in an MRI environment.

The downsides of this technique mostly pertain to its spatial resolution. Infrared light can travel approximately 1-2cm below the scalp, which corresponds to less than 10mm of cortex before scattering (Scholkmann et al., 2014). Even so, prior studies show a good correspondence between an fMRI-derived BOLD signal and fNIRS (Ferradal et al., 2014). Recent work, for instance, tested an auditory sentence comprehension in both fMRI and fNIRS to show similar findings of a hierarchical speech network as well (Hassanpour et al., 2015).

Although the first fNIRS study of speech processing was in 1998 with stroke patients (Sakatani et al., 1998), only in recent years has it begun to be applied to CI studies [for a review: (Bortfeld, 2019)]. The first fNIRS study of pediatric CI users, for instance, was published less than a decade ago (Sevy et al., 2010). In total, there are now over a dozen fNIRS studies of CI users, all of which further validate this technique as both safe and effective in this patient population.

1.12 An overview of the experiments and hypotheses in subsequent chapters

This dissertation is framed from the perspective that visual orofacial articulations play a crucial role in verbal communication both before and after cochlear implantation, and in order to fully describe aural speech recovery following implant surgery, characterization of both unisensory and multisensory processing is necessary. Importantly, it is currently unclear whether visual or audiovisual performance varies as widely as auditory-only speech measures. Also, despite substantial evidence for changes in visual functional capabilities during deafness, their relationship to outcomes following cochlear implantation has not been characterized. We believe that this is a critical piece in understanding functional outcomes, as speech (and various other facets of naturalistic environments) is inherently audiovisual. Hence, differences in visual abilities may play an underappreciated role in implant outcomes. The studies herein utilize a dozen different visual and audiovisual tasks to test a range of questions on this topic (Table 1.1).

CH	Sample	Age range (y)	Question	Tasks
2	CI n = 63 NH n = 69	6-77	Do CI users experience AV illusions differently from NH controls? If so, how do these measures relate to clinical outcomes?	• McGurk illusion • Sound-induced flash illusion • Clinical metrics
3	NH n = 24	19-34	Can we use fNIRS to localize cortical activity in response to AV speech?	• McGurk illusion • AV word recognition in noise • fNIRS (52 channel)
4	CI n = 23 NH n = 23	33-71	Do the behavioral and neuroimaging experiments we designed in chapter 3 reveal differences in AV speech processing between CI users and age-matched NH controls?	• McGurk Illusion • AV word recognition in noise • fNIRS (24 channel) • Clinical metrics
5	CI n = 48 NH n = 54	19-77	Are there underlying differences in temporal processing between CI users and controls? If so, are they specific to speech or involve non-speech stimuli as well?	6 psychophysics tasks of: • Simultaneity judgment • Temporal order judgment

Table 1.1. Overview of all participants and studies in the following four data chapters. AV = audiovisual, fNIRS= functional near infrared spectroscopy, CI = cochlear implant, NH = normal hearing.

The overarching goal of this body of work is to quantify the role of vision in hearing restoration through cochlear implants. Doing so involves: the pairing of incongruent auditory and visual information to elicit AV illusions (Chapter 2), the development and testing of a novel paradigm for behavioral AV speech integration and neuroimaging in NH controls (Chapter 3), between-group comparisons with CI users and age-matched controls on these tasks (Chapter 4), and finally, an investigation of the temporal judgments of auditory and visual stimuli—a determining factor for whether they are integrated into a multisensory percept (Chapter 5).

1.13 Potential broader impact of this work

As the criteria for CI candidacy broaden (Deggouj et al., 2007), there is a growing need to both refine postoperative measures of speech proficiency and to identify additional factors that influence remaining variability in what is now a large clinical population. For instance, it is unclear whether CI users with poor auditory speech understanding actually display more AV gain than their peers through the addition of visual cues. If so, current auditory-verbal therapeutic approaches that actively discourage the utilization of visual speech cues, may be overlooking greater perceptual advantages of multisensory integration for aural rehabilitation in everyday listening environments. Given that even short periods of sensory deprivation have been linked to compensatory plasticity (Pascual-Leone and Hamilton, 2001), visual proficiency may be an understudied avenue for further remediation. In contrast, if AV integration is hindered in poor auditory-only performers, then perhaps pre-operative visual abilities (and the associated changes in brain circuits) may predict this difficulty. Each of these potential outcomes has significant clinical value for better predicting outcomes, and, therefore, managing patient expectations prior to implantation. This knowledge is essential for our understanding of speech proficiency with a CI and, most importantly, for how users can best utilize all sensory information to enhance speech comprehension and improve quality of life.

1.14 Acknowledgements

Figures 1.3 and 1.4 appear in the following manuscript:
Chern, A., Butera, I.M. Multidisciplinary perspectives on music perception and cognition in cochlear implant users. *Music and Medicine*. In press.

CHAPTER 2 Visually-biased illusion perception in cochlear implant users

2.1 Introduction

Speech, like many other daily activities, is typically multisensory. For cochlear implant (CI) users, talking on the phone is a difficult—and oftentimes insurmountable—challenge. Thus, for many users, audiovisual conversations are a necessity. Despite this daily reliance on the multisensory integration of speech, audiovisual testing is not a part of routine audiological exams. Furthermore, as a clinical population, there are substantial individual differences (e.g., type and onset of hearing loss, duration of deafness, and age of implantation) that contribute to variable auditory outcome measures and may also have cascading effects on AV integration (Blamey et al., 2013). Arguably, the index of multisensory processing that we know the most about in CI users is the McGurk illusion (McGurk and MacDonald, 1976). First described over 40 years ago, this illusion results from hearing a bilabial syllable (e.g /ba/ or /pa/) while seeing the visually-ambiguous articulation of a velar syllable (/ga/ or /ka/). Together, these elicit a novel, fused percept such as /da/, /tha/, or /ta/. This effect is both robust and persistent for individuals who experience it, though not all people do.

There may be as few as 25% of people who don’t experience the illusion or as many as 75%, depending on the experimental design (e.g., trial order, open set v. forced choice responses, and the auditory and visual stimuli themselves) (Mallick et al., 2015). Many studies of CI users have focused on the responses of these “non-perceivers,” because it provides insight into sensory biases when faced with conflicting audiovisual information. There is remarkable consistency in all studies evaluating the McGurk illusion in CI users that those who do not perceive the illusion are biased toward the visual component (see Table 2.1 for brief summaries and citations). This finding is in contrast to NH controls who typically report the auditory component when not experiencing the illusion (Massaro et al., 1986; McGurk and MacDonald, 1976). This discrepancy makes intuitive sense, because many CI users struggle to discriminate auditory-only syllables and may be perceptually “weighting” visual information more highly in order to improve AV estimates. Indeed, adding noise or otherwise altering the saliency of one sensory modality is well-known to effectively and rapidly alter sensory weights in typical populations as well.

Despite this consensus in the literature regarding visual bias among CI users, the rate of fusion relative to NH controls is less apparent. Some studies indicate similar rates (Huyse et al., 2013;

Rouger et al., 2008; Tremblay et al., 2010), while others find CI fusion that's both lower (Schorr et al., 2005) and higher than controls (Desai et al., 2008a; Stropahl et al., 2017b; Tona et al., 2015). Given the more unreliable auditory percepts among CI users, comparisons to NH individuals requires some derivation. Take, for example, an auditory error in unisensory trials like mistaking the sound of /ba/ for /da/. Because this could mimic a McGurk percept in the incongruent trials, an adjustment is necessary to better quantify illusion perception *per se* (Grant et al., 1998; Stevenson et al., 2012). This need may be particularly true for patient populations with hearing impairments. Desai and colleagues, for instance, found much higher fusion response rates in CI users (60% v. 20%) that were actually not significantly different after applying an error correction (2008a). Further investigations are needed to address this issue, particularly in larger sample sizes that also capture the clinical diversity of CI users today.

Citation	Study size	Study sample (ages)	McGurk Stimuli*		Key findings
			A	V	
Schorr et al. 2005	CI n = 36 NH n = 35	Children (5-14y)	pa	ka	Children with CIs are more likely to fuse syllables if they receive their implant prior to the age of 2.5y, otherwise they perceive the visual syllable /ka/.
Rouger et al. 2008	CI n = 33 NH n = 39	Adults CI_{avg}= 52y NH_{avg}=32y	aba apa ama	aga	CI bias toward the visual information, though illusion perception is similar to controls.
Desai et al. 2008	CI n = 8 NH n = 14	Adults (18-80y)	ba	ga	CI users rely more on the visual cue and their McGurk fusion rate positively correlates to duration of implant experience.
Trembley et al. 2010	CI n = 17 NH n = 12	Adults (19-69y)	ba	ga	Proficient CI users, like NH controls, responded with either /da/ or /ba/, while non-proficient CI users are more visually biased.
Huyse et al. 2013	CI n = 31 NH n = 31	Children (7-17y)	afa apa	asha aka	Non-proficient CI kids' responses were more visual and less auditory than proficient kids.
Tona et al. 2015	CI n = 24 NH n = 12	Children (4-10y)	ba pa	ga ta	The McGurk illusion was experienced more often in Japanese children with CIs than age-matched controls.
Stropahl et al. 2016	CI n = 8 NH n = 24	Adults (19-75y)	ba ma pa	ga ka	CI users have higher sensory noise and a higher threshold of perceiving the illusion compared to controls in a stimulus-independent NED model.
Yamamoto et al. 2017	CI n = 31	Children (5.5-14y)	ba pa	ga ka	Prelingually deafened Japanese children with bilateral CIs experienced the illusion while listening with just one or with both implants.

Table 2.1. McGurk studies in the literature. All studies report a visual bias in CI users, while Schorr, Desai, and Tremblay also found various correlations to clinical measures.

*Only auditory bilabial stimuli and visual velar stimuli are listed, though it should be noted that four studies (Rouger, Stropahl, Tona, and Yamamoto) also tested several other syllables in up to 12 different combinations. NED = noisy encoding of disparity (Magnotti and Beauchamp, 2015).

In addition to this outstanding question of whether CI McGurk perception differs from NH listeners, it is also unknown if these results directly relate to other metrics of AV integration. The sound-induced flash illusion—sometimes referred to as the flashbeep illusion—was discovered more recently and involves simple, non-speech stimuli (Shams et al., 2000). Unlike how the McGurk task asks participants what they heard, the sound-induced flash illusion, or SIFI, asks participants to ignore what they hear and only report what they saw. Specifically, they are to count the number of rings rapidly flashing on a screen while they also hear multiple beeps. The more the added beeps, the more likely multiple flashes are perceived. Results from this task would be particularly interesting in CI users given that deafness has been linked to enhancements in the visual periphery (Bavelier et al., 2006), and this illusion is perceived more strongly in the parafoveal visual fields of typical listeners (Shams et al., 2002). We hypothesized that the aforementioned visual biases in CI users lead to more veridical visual perceptions, which would correspond to a *decreased* likelihood of perceiving the non-existent flashes in a SIFI task. Furthermore, because low McGurk fusion is effectively a metric of visual bias for CI users, we predicted a positive correlation between the two tasks. That is, when simple visual percepts are more easily biased by sound, we anticipate a higher fusion of auditory and visual speech as well.

This study tests the McGurk illusion in the largest CI cohort to-date and is the first to investigate the Sound-induced flash illusion (SIFI) in this population. Our aim is to investigate how these tasks relate to one another and to clinical outcome measures for CI users. Such work is necessary to better characterize AV integration in this cohort of individuals for whom it may be exceedingly important to “bind” auditory and visual information into more reliable multisensory percepts.

2.2 Methods

Participants. We recruited 63 CI users and 69 NH controls (Table 2.2). There was no significant difference between groups for age ($t_{(130)} = 1.9$, $p = 0.06$). At least 3 months of experience with their implants was an inclusion criterion for CI users, and the average was 42 months post activation (range = 4 mo – 11 y). The duration of hearing impairment (i.e., from the known onset of hearing loss) was 23.6 years on average (range = 1.5-73 y). We also calculated an approximate “duration of deafness,” which we defined as the date of cochlear implantation minus the onset of severe hearing loss. We used a standard definition of severe hearing loss as either the date when pure tone detection exceeded 70dB or else the closest estimate from clinical records. In this cohort, the average duration of deafness is 2.4 years (range 0 – 27 y). The majority of CI users (n = 53) were postlingually deafened; however, a subset was prelingually deafened (n = 10). These individuals received their implants at an average age of 3.5 years (range = 1.3–6.5y; see black circles in Fig. 2.1). Although prelingual deafness (i.e., occurring prior to language acquisition and during critical periods of development) can negatively impact speech outcomes (Fitzpatrick, 2015; Miyamoto et al., 1994), all of these 10 individuals have high speech proficiency (Fig. 2.1). As a result, we included all 63 CI users in the analyses.

Group	N	Sex (% female)	Mean age ± SD (y)	Number of CIs	Acoustic hearing	Implant manufacturer	Duration of hearing impairment (y)
CI	63	56%	44.6± 21.1	one – 52% two – 48%	41%	60% Cochlear 25% MED-EL 15% AB	23.6 ± 17.4
NH	69	73%	37.8± 20.1	—	100%	—	—

TABLE 2.2. Participant characterization. A large cohort of cochlear implant (CI) and normal hearing (NH) controls ranging in age from 6 to 77 years old participated in this study. Nearly half of the CI users (48%) had bilateral implants while many of the others had residual acoustic hearing (41%). Among the 93 implanted ears in this study, the majority (60%) were manufactured by Cochlear, Ltd, followed by Med-El and Advanced Bionics—a typical distribution for US patient populations.

All CI users completed testing in their "best-aided" hearing condition, which included hearing aids for 41%. All participants wore corrective lenses as needed and were screened for visual acuity using either a Snellen eye chart or verbal report. Speech perception in the CI group was either tested at the study visit or recorded from recent medical records. This measure was not available for three individuals, and of the remaining 60 CI users, one only had LNT testing, while the rest are best-aided, monosyllabic CNC word scores out of a possible 100% correct. Results indicate a wide range of proficiency from 0%-98% correct (Fig. 2.1), and this variability is consistent with other reports in the literature (Gifford et al., 2008).

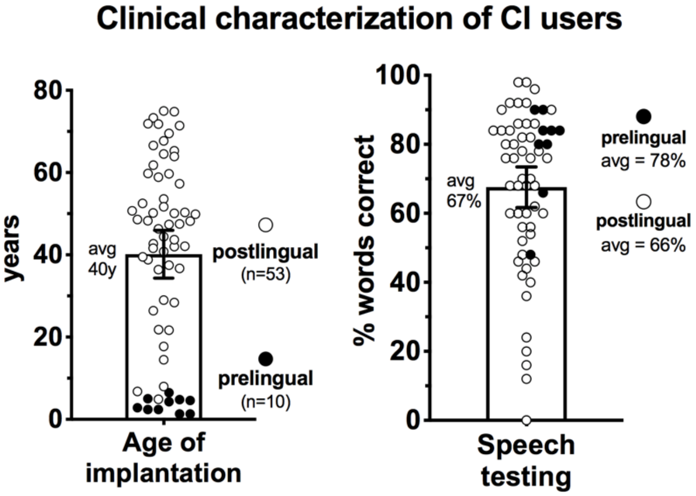


FIGURE 2.1. Clinical outcomes for the CI group. Age of implantation ranged from 1.3-75 years old (left), and all users had at least 3 months of CI use prior to participation. Clinical speech testing had a wide range of outcomes (right). We see higher than average performance in prelingually deafened individuals (filled circles) who were all implanted very young. Bars indicate group means and error bars are 95% confidence intervals of the mean.

Stimuli. Visual stimuli are displayed using Matlab 2008a and Psychophysics toolbox extensions (Brainard, 1997). These stimuli are presented on a CRT monitor positioned approximately 50cm from participants. Visual stimuli are white circles (13 ms in duration) on a black background and 2s videos of a female articulating the syllables /ba/ and /ga/ (Stevenson et al., 2014a). Auditory stimuli are 3.5 kHz tones (50 ms in duration) in the SIFI task and utterance of the syllables /ba/ and /ga/ in the McGurk task. These auditory stimuli are delivered at a comfortably loud level (approximately 65 dB SPL) through a mono speaker. The aligned onset of visual and auditory stimuli is confirmed using a Hameg 507 oscilloscope via inputs from a photovoltaic cell and a microphone.

Task	Total trials	Stimuli		Illusion	Prompt
		Visual	Auditory		
McGurk illusion	88	 "ga"	 	 	Report the first letter of the syllable the speaker said using the letters b, g, d, or t.
Sound-induced flash illusion	275				Ignore the beeps, and use the number pad to report the number of flashes you saw: 1, 2, 3 or 4.

FIGURE 2.2. Experiment details of the McGurk illusion and the sound-induced flash illusion (SIFI). These two tasks have both congruent control trials and incongruent illusory trials where participants are asked to either report what they heard (McGurk) or what they saw (SIFI).

Experimental Design and Analysis. The McGurk task has two blocks. The first block is unisensory testing of auditory-only and visual-only presentations of /ba/ and /ga/ (28 trials total). The next block is all audiovisual presentations of both syllables (40 trials) and the incongruent pairing of auditory /ba/ with the visual articulations of /ga/ (20 trials). Participants respond to each trial with one of four letters corresponding to the sounds: /ba/, / ga/, /da/, or /tha/ (Fig. 2.2). Illusory responses were considered either /da/ or /tha/, which we'll simply refer to as /da/ henceforth. Because misidentifying unisensory stimuli as one of these novel syllables could overinflate the apparent magnitude of the illusion, we made a correction using the formula:

$$p(\text{“da”} \mid \text{McGurk trial}) \times [1 - p(\text{“da”} \mid \text{Unisensory trial})]$$

Thus, subtracting /da/ responses in unisensory trials from /da/ responses to McGurk trials effectively lowers the probability of experiencing the illusion—p(McGurk)—for those who struggle to distinguish the component syllables on their own.

In the SIFI task, participants fixated on a white cross in the middle of the screen while a white ring flashed on a back background (Fig. 2.2). They were asked to report the number of flashes while ignoring the beeps. Control trials consisted of: 1-4 flashes without beeps, congruent numbers of flashes and beeps (up to 4), and incongruent pairings of just one flash with 2-4 beeps. 25 trials were tested for each of these 11 conditions, and responses were scored as the average number of flashes reported. Additionally, we calculated a susceptibility index (SI) using the following formula where R_n is the average number of flashes for n beeps:

$$SI = \frac{\frac{R_2 - R_1}{1} + \frac{R_3 - R_1}{2} + \frac{R_4 - R_1}{3}}{3}$$

This index essentially collapses across all incongruent trials in order to derive a single metric for the average number of illusory flashes that are experienced per added beep. A value of 0 means that no illusion was perceived, and a value of 1, for instance, indicates that each additional beep increased the perceived number of flashes by 1.

Procedures. All protocols and procedures were approved by Vanderbilt University Medical Center's Institutional Review Board, and all subjects provided informed consent prior to participation. Experiments took place in a dimly lit, sound-attenuated room with an experimenter seated nearby. Both task order and trial order were pseudo randomized. On all non-speech tasks, subjects were instructed to maintain fixation on the centrally-located fixation cross. All responses were collected using a standard keyboard. These experiments were part of a larger testing battery (Butera et al., 2018). Trial and task order was pseudorandomized. Because some individuals ran out of time at the end of testing, subject numbers are included on figures for those who were able to complete the task.

Statistical approach. Between-group differences in the McGurk task were tested using resampling methods on account of several highly skewed variables that are incompatible with parametric tests even after standard transformations (e.g., $\log_{10}$). This approach was selected for its minimal assumptions of the data's distribution in each group, and is based on 30,000 reshuffles where each shuffle reassigns data points randomly to the two groups. A two-tailed, Welch's t-test is then used as a comparison metric. We selected this metric instead of a difference in means, for instance, because it has the advantage of capturing both a central tendency and variability. Significance is expressed as the number of times a shuffled sample produces a p value exceeding what was found in the actual sample, and is expressed as a proportion of the total number of simulations. T scores and degrees of freedom are reported from the Welch's t-test of the observed data. Control McGurk trials, the p(McGurk) index, and SI were compared between groups using this resampling method in R.

Data from the SIFI task was less skewed and a square root transformation corrected for small deviations from normality. Congruent and incongruent conditions were both compared via a mixed model ANOVA in SPSS. The repeated measure is the number of added beeps, and the two groups are the between-subjects factor. Greenhouse-Geisser corrections for violations of sphericity are applied as needed, and significant interactions are followed up with pairwise t-tests.

Correlations between tasks metrics were done with a Pearson's correlation in SPSS software. Lastly, for a linear regression within the CI group, we included uncorrelated clinical variables with log transformations as needed to correct for deviations from normality. Significance for all statistical tests was defined as $\alpha < 0.05$.

2.3 Results

McGurk Illusion. In both the unisensory and AV control trials, CI users have lower speech perception accuracy for both /ba/ and /ga/ stimuli (Table 2.3). Not surprisingly, the group differences are largest for the auditory-only conditions (Fig. 2.3a). Additionally, for identifying /ga/, CI users had lower accuracy in lipreading (CI mean= 53%, NH mean = 68%) and AV listening (CI mean= 83%, NH mean = 99%). Though AV identification of /ba/ was also statistically lower for CI users, both groups scored close to ceiling (CI mean =97.7% and NH mean= 99.3%).

Modality	Syllable	t statistic	Significance
Auditory-only	Ba	$t_{(92.2)} = -9.6$	**p < 0.001**
	Ga	$t_{(62.3)} = -6.8$	**p < 0.001**
Visual-only	Ba	$t_{(94.3)} = -1.4$	p = 0.16
	Ga	$t_{(125.9)} = -3.0$	**p = 0.004**
Audiovisual (congruent)	Ba	$t_{(74.1)} = -2.3$	**p = 0.02**
	Ga	$t_{(63.4)} = -6.7$	**p < 0.001**
Audiovisual (McGurk)	Ba	$t_{(70.5)} = -5.9$	**p < 0.001**
	Ga	$t_{(119.4)} = 6.0$	**p < 0.001**
	Da	$t_{(121.0)} = 0.15$	p = 0.87

TABLE 2.3. McGurk results. Control trials and incongruent AV "McGurk" trials are compared between groups using a resampling approach. Significant values (p< 0.05) are bolded.

In the incongruent McGurk trials (Fig. 2.3b), there is no significant difference in /da/ responses (Table 2.3). However, when CI users did not fuse the syllables, they were much more likely to report the visual component /ga/, Conversely, non-fusing NH controls were most likely to report the auditory component /ba/. Individual data from these McGurk trials illustrate these biases as well as the high proportion of NH controls who did not perceive the illusion (Fig. 2.3c, white circles). In NH populations, prior studies have described the McGurk illusion occurring as an "all-or-nothing" effect such that the majority of individuals either experience the illusion almost always or else rarely ever (Mallick et al., 2015). In the present study, we found that 72% of the NH group fell within the two extremes of perceiving the illusion either ≥90% of the time or ≤10%. In contrast, the CI group had many more intermediate responses with only 35% of individuals having very high or very low fused responses (see /da/ panel in Fig. 2.3c).

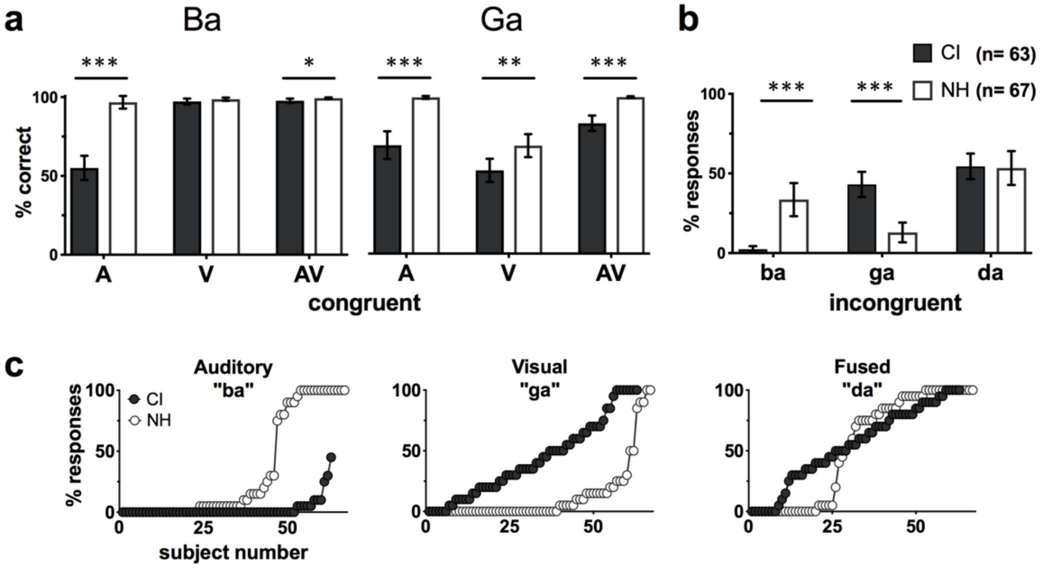


FIGURE 2.3. McGurk experiment results. Mean accuracy of perceiving syllables in both unisensory and congruent AV trials are shown for /ba/ and /ga/ (a). Mean responses to incongruent "McGurk" trials are shown in (b), and individual data to each response is shown in (c). Error bars indicate 95% confidence interval of the mean. * p< 0.05, **p< 0.01, ***p< 0.001.

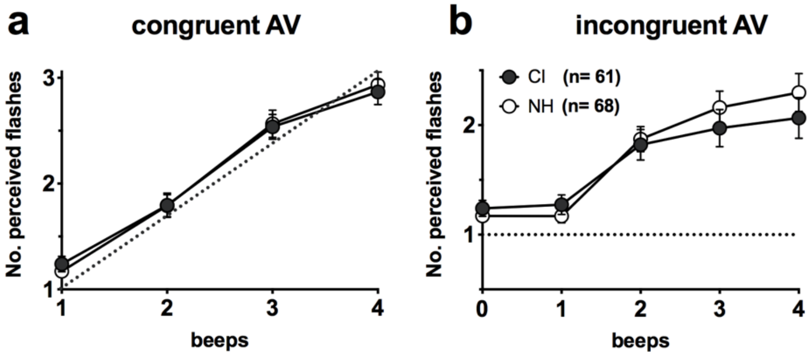


FIGURE 2.4. Sound-induced Flash Illusion results. Average reports of the number of perceived flashes are plotted for each group. Trials were either (a) congruent pairings of one or more flashes *and* beeps or (b) 0-4 beeps paired with only one flash. Dotted lines represent 100% accuracy.

SIFI. In control trials where the number of flashes and beeps is matched (Fig. 2.4a), there are no between-group differences in the perceived flashes ($F_{(3,127)} = 0.002$, $p = 0.97$, $\eta_{partial}^2 = 0$). In the incongruent pairings of just one flash with multiple beeps (Fig. 2.4b), there is also no between-subjects effect ($F_{(4,127)} = 0.55$, $p = 0.46$, $\eta_{partial}^2 = 0.004$). However, there is a beep × group interaction ($F_{(1.4,127)} = 7.6$, $p = 0.003$, $\eta_{partial}^2 = 0.056$). Follow-up tests did not identify any group differences at specific conditions, though several approached a p value of 0.05 (Table 2.4). Most notably, CI users reported slightly fewer flashes in the three-beep (avg flashes = 1.96) and four-beep conditions (avg flashes = 2.07) compared to controls (2.16 and 2.30, respectively).

Condition	t statistic	Significance
1 flash	$t_{(127)} = 1.5$	$p = 0.13$
1 flash + 1 beep	$t_{(112.2)} = 2.0$*	$p = 0.051$
1 flash + 2 beep	$t_{(127)} = -0.69$	$p = 0.49$
1 flash + 3 beep	$t_{(127)} = -1.8$	$p = 0.081$
1 flash + 4 beep	$t_{(127)} = -1.9$	$p = 0.066$

TABLE 2.4. SIFI results. Though within-subjects effects from a mixed-model ANOVA indicted a significant beep × group interaction, no follow-up t-tests reached the p = 0.05 significance threshold. *Fractional degrees of freedom result from unequal variances as identified by a Levene's test.

Illusion metrics. In order to compare group differences in illusion perception between these two tasks, we calculated p(McGurk) and SI metrics (Fig. 2.5). For p(McGurk) there is no difference between CI and NH groups ($t_{(109.6)} = -1.7$, $p = 0.097$). However, we also compared whether the magnitude of the illusion differed strictly among individual who did perceive the illusion in at least one trial. Thus, we excluded all "non-perceivers" who made up 13% of the CI group (n = 8) and 30% of the NH (n = 20). Of the remaining 55 CI users and 47 NH controls, we see a highly significant difference between p(McGurk) measures ($t_{(81.0)} = -4.9$, $p = 4 \times 10^{-5}$). Following the same trend, there is also significantly lower SIFI susceptibility for CI users ($t_{(124.5)} = -2.5$, $p = 0.014$; Fig. 2.5b).

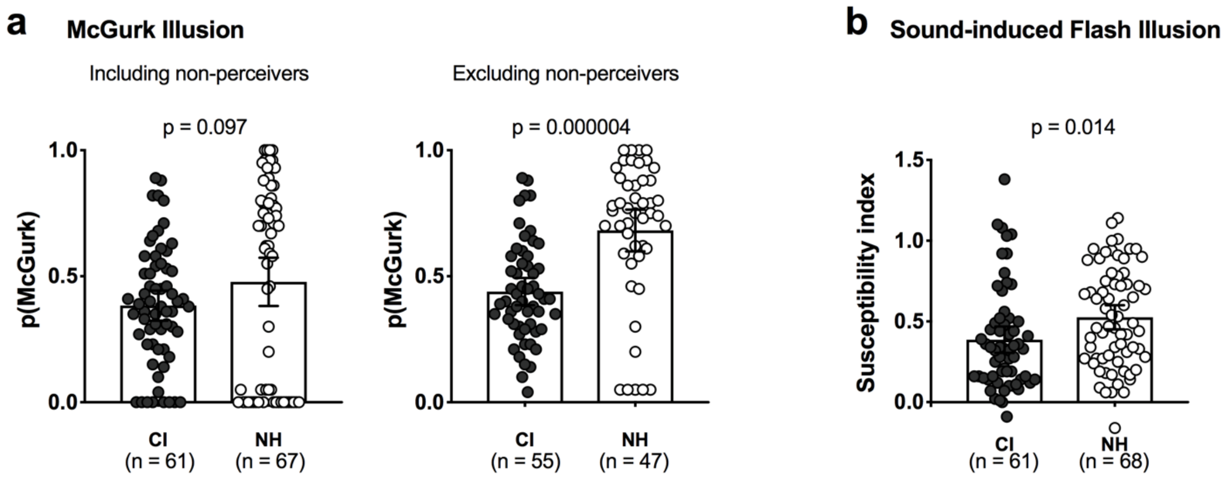


FIGURE 2.5. Derived metrics of the probability of McGurk perception and SIFI susceptibility index. The p(McGurk) metric corrects for inaccurate "da" percepts in control conditions (i.e., which only contain "ba" and "ga"). Susceptibility index collapses across all incongruent conditions (i.e., 2 to 4 beeps per flash) to quantify the average increase in perceived flashes per added beep.

Correlation between illusion tasks. Next, we asked whether these two illusions have a relationship to one another in these cohorts. Neither Pearson's correlation met the significance threshold; however, a negative relationship in the NH group did approach significance (p = 0.063; Fig. 2.6b).

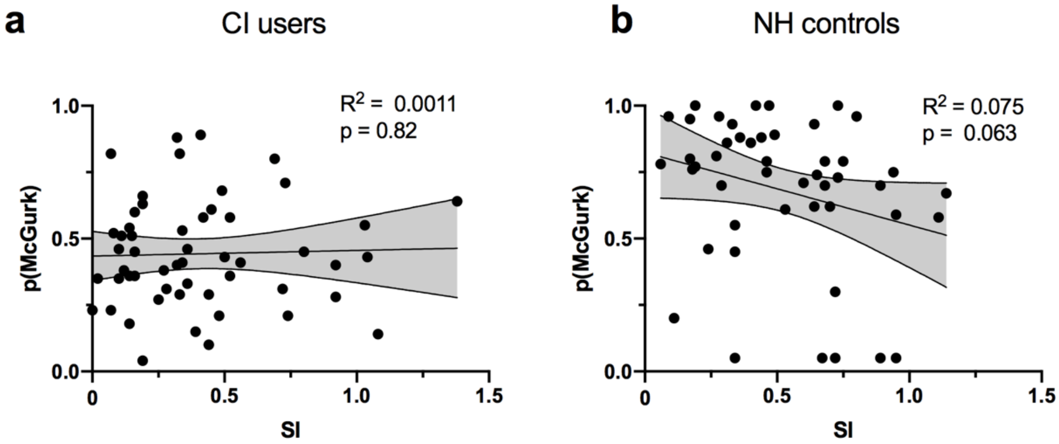


FIGURE 2.6. Relationship between SIFI and McGurk. The correlation between susceptibility index (SI) and probability of McGurk perception is not significant for CI users (a) or NH controls (b). Shaded areas correspond to 95% confidence intervals.

Testing for additional explained variability in clinical measures. Lastly, we asked whether these AV illusion metrics have any predictive value for explaining variability in clinical speech scores within the CI group (Fig. 2.1). In a stepwise regression model with CNC scores as the dependent variable, we entered the following four uncorrelated clinical and experimental measures as independent variables: duration of hearing impairment, duration of deafness, p(McGurk), and SI. The only variable that is a significant predictor of CNC scores is the duration of hearing impairment, explaining 12.9% of variability ($R^2 = 0.129$, $F_{(1,41)} = 7.2$, $p = 0.01$). All other variables were excluded from the model.

2.4 Discussion

This study tested the McGurk illusion in the largest CI cohort to-date and is the first to investigate the Sound-induced flash illusion (SIFI) in this population. A key finding in this study is that CI users perceived both of these AV illusions less often than NH controls. Additionally, we replicated the same general trend as others have reported in the McGurk task: that CI users display a visual speech bias. Importantly, our results also illustrate the need to correct for mistaking unisensory syllables as /da/. That is, in this study, like several others (Huyse et al., 2013; Rouger et al., 2008; Tremblay et al., 2010), /da/ responses were the same between groups (Fig. 2.3c). It was only in comparing the p(McGurk) measure of individuals who perceived the illusion at least once that we saw the groups diverge (Fig. 2.5). Based on these findings, we recommend that any further investigations of this illusion in CI users take similar measures to disambiguate /da/ responses and fused percepts.

Similarly, in the SIFI task, CI users did perceive the illusion, but to a lesser extent than controls. We saw small deviations in several flash-to-beep ratios (Table 2.4) that, after collapsing across all conditions into a susceptibility metric, was significantly lower than controls (Fig. 2.5b). This novel finding is noteworthy given that relatively little is known about low-level stimulus detection in CI users (Stevenson et al., 2017a). Based on prior work from our lab, we do know, however, that CI users' temporal judgments of synchrony for these same flashbeep stimuli are indistinguishable from controls (Butera et al., 2018). This suggests that low-level AV temporal function is likely intact in postlingually-deafened adult, so any broader issues, like the SIFI magnitude itself, may be subtle. Although work in other clinical populations has related reduced SIFI perception to broader impairments in integration (Stevenson et al., 2014a), further work is needed to better characterize any further implications and mechanisms in CI users.

It should be noted that the illusion of multiple flashes can be elicited in other modalities as well (Violentyev et al., 2005). In an fMRI study of deaf individuals, a visual-tactile illusion analogous to SIFI, revealed larger visual responses in Heschl's gyrus and a greater magnitude of the illusion experienced in deafness (Karns et al., 2012). Given what is already known about the mechanisms behind the sound-induced flash illusion (Cecere et al., 2015; Kerlin and Shapiro, 2015; Watkins et al., 2006), it would be interesting to test whether, for instance, visually-derived alpha oscillations play a similar role in predicting the likelihood of experiencing the illusion in CI users or if other, crossmodal mechanisms are at play.

A broad interpretation of the present study is that CI users perceptually "weight" visual information more highly, which disrupts both McGurk and SIFI illusions. That is, in incongruent tasks when the "correct" answer is visual (i.e., SIFI), they have more veridical percepts. However, when the correct answer is auditory (McGurk), then they are less accurate and more visually biased. Thus, in both cases of incongruency between the senses, visual input is prioritized. A similar pattern of more visually-biased McGurk responses can also be simulated in NH users by vocoding stimuli to sound like cochlear implants (Desai et al., 2008a). Similarly, adding a visual distractor reduces McGurk percepts (Tiippana et al., 2004) and so does adding noise to reduce the auditory and/or video salience (Fixmer and Hawkins, 1998; Hirst et al., 2018). Further work is needed to understand the clinical impacts of this visual bias in CI users. Though prior studies have suggested that more proficient CI users experience greater fusion (and less visual bias) (Huyse et al., 2013; Tremblay et al., 2010), we did not find McGurk fusion to explain any greater variability in CNC scores than other known measures—in our case, the total duration of hearing loss.

Additionally, we did not see a direct relationship between SI and p(McGurk) in CI users as we had anticipated (Fig. 2.6), though there was a correlational trend in the NH group ($R^2 = 0.075$, $p = 0.063$). In a slightly younger and smaller NH sample size, Stevenson and colleagues (2012) reported a stronger, also negative correlation between p(McGurk) and SIFI ($R^2=0.42$ $p = 0.003$). In contrast, Tremblay et al. (2007) found no correlation between SIFI and McGurk tasks in a sample of 38 typical listeners between 5 and 19 years of age. While it is logical that differences in low-level AV illusions indexed by SIFI could have cascading effects on speech integration, these two tasks do not appear to have a direct relationship in CI users.

One caveat to our study is that that McGurk perception is highly stimulus-specific, and we only tested responses to one female speaker. As a result, it is possible that these results may not generalize more broadly to other syllable combinations, male speakers, languages, etc. The primary advantage for interpreting our findings is that we tested a large sample of clinically-diverse CI users, so CI perception to this stimulus is well-characterized. In future studies, applying a noisy encoding of disparity (NED) model to quantify results would be ideal for future studies investigating the effect in a stimulus-independent manner (Magnotti and Beauchamp, 2015; Stropahl et al., 2017b).

Finally, it is not entirely clear what these results practically mean for cochlear implant outcomes given the lack of a relationship with CNC scores. Are more visually-biased CI users less proficient with their implants? Are they worse "multisensory integrators"? Because it's more likely that this AV illusion would relate to audiovisual speech measures than the auditory-only ones tested in clinic (and reported here), future work comparing illusion perception to a real-world estimate of one's success in the integration of conversational speech would provide more insight. Until we know how McGurk biases relate to natural speech integration of words or sentences it is unclear what utility, if any, administering McGurk testing at a broader scale might have. Either way, better identifying individual differences in multisensory integration is an important next step, particularly for the development of new audiovisual remediation strategies.

2.5 Acknowledgments

Ryan Stevenson designed the experiments and collected data. Brannon Mangus recruited participants. Iliza Butera collected and analyzed the data and wrote the manuscript. Dan Ashmead wrote an R script for the resampling analysis. Mark Wallace and René Gifford provided guidance and feedback throughout the experiments and manuscript preparation.

CHAPTER 3 Functional localization of audiovisual speech using near infrared spectroscopy

3.1 Introduction

The “cocktail party effect” describes a common circumstance where attention must be focused on a single conversation while in the presence of many others. The fact that the background noise, or babble in this case, is most effectively “filtered out” with the help of lipreading has been known for some time (Sumby and Pollack, 1954). Even so, there are no standard audiovisual tasks of this effect in audiological test batteries. In this study, we sought to design a cocktail party task that was comparable to other standard clinical speech testing in the auditory-only domain but that also included visual-only and audiovisual speech recognition for a range of possible signal-to-noise ratios (SNRs).

The superior temporal sulcus (STS) is a known locale for the integration of AV speech, and the organization of auditory, visual, and AV inputs to STS has been described for some time (Beauchamp et al., 2004). However, limitations in available imaging techniques have been a barrier for similar functional measurements in cochlear implant (CI) users. The field of functional Near Infrared Spectroscopy (fNIRS) recently celebrated its 20th anniversary (Ferrari and Quaresima, 2012), and this technique has begun to be applied to CI users (Saliba et al., 2016). This rapidly-growing body of work supports: the sensitivity of fNIRS for measuring speech-evoked activity in NH controls (Pollonini et al., 2014) and CI users (Sevy et al., 2010), distinguishing these activity patterns between proficient and non-proficient CI users (Olds et al., 2015), distinguishing unique phonological awareness networks in non-proficient CI users (Bisconti et al., 2016), and measuring multisensory interactions for low level stimuli (Wiggins and Hartley, 2015) and speech stimuli (Anderson et al., 2017; van de Rijt et al., 2016; Stropahl and Debener, 2017). As a natural extension of this current area of inquiry, we sought to first characterize neural processes underlying AV integration of monosyllabic words in a normal-hearing cohort. Following this technical validation, later comparisons could be made to age-matched CI users. Thus, by measuring optically-derived hemodynamic responses, we can safely and non-invasively investigate cortical activation in NH controls and CI users.

Because maximum STS recruitment is known to be elicited by AV speech when presented in noise (Callan et al., 2003), we test two signal-to-noise ratios (SNRs), chosen from pilot data to find moderately-difficult noise levels that elicit high multisensory gain (Ross et al., 2007). Behavioral and neuroimaging measures of AV word recognition are also compared to the likelihood of perceiving the McGurk illusion. Garnering over 6,000 citations for the original paper, the McGurk illusion has also been tested in many clinical populations; however, recent work in normal hearing populations questions the utility of this measure (Van Engen et al., 2017) and whether it may actually rely on distinct processing pathways from the congruent AV speech perception that, historically, it has been presumed to index (Erickson et al., 2014; Hickok et al., 2018).

In summary, the aim of this study is two-fold: first to design and test behavioral and neuroimaging tasks of AV integration, while also comparing the results of word recognition in noise to the McGurk effect in a normal hearing population.

3.2 Methods

Participants. We recruited 24 adults with normal hearing to participate in a 2h study of behavioral speech testing and optical neuroimaging. All participants were native English speakers and passed hearing screenings (i.e., detection levels no greater than 20 dB HL) at 250, 500, 1k, 2k, 4k Hz frequencies. The average age was 25.4 years (SD= 5.1), and 45% were female (n = 11). Twenty-one participants (88%) were right handed. All procedures were approved by the Vanderbilt University Institutional Review Board, and all individuals provided written informed consent. Data was collected from April to May 2018, and participants were compensated with a $30 gift card upon completion of the study.

fNIRS. Neuroimaging was done using NIRScout equipment (on loan from NIRx). A cap containing 16 LED sources and 23 detectors was aligned with the 10-20 points shown in Fig 3.1a. This arrangement resulted in 52 recording channels that were divided into 4 ROIs for analysis. Data was collected in an interleaved manner at a sampling rate of 7.8 Hz. The concurrent task involved word categorization in 4 different conditions: auditory only listening in noise, audiovisual listening in noise, visual-only lipreading, and reading written words (Fig. 3.1b). These conditions were each convolved with hemodynamic response functions within each 20s “ON” block to create a model of the change in oxygenated hemoglobin (oxyHb) concentration (Fig. 3.1c). Participants responded with a keypad to indicate in which of two categories a word belonged (objects v. numbers or actions v. animals; see Appendix A for full word lists). Data from two runs were averaged together in the analysis.

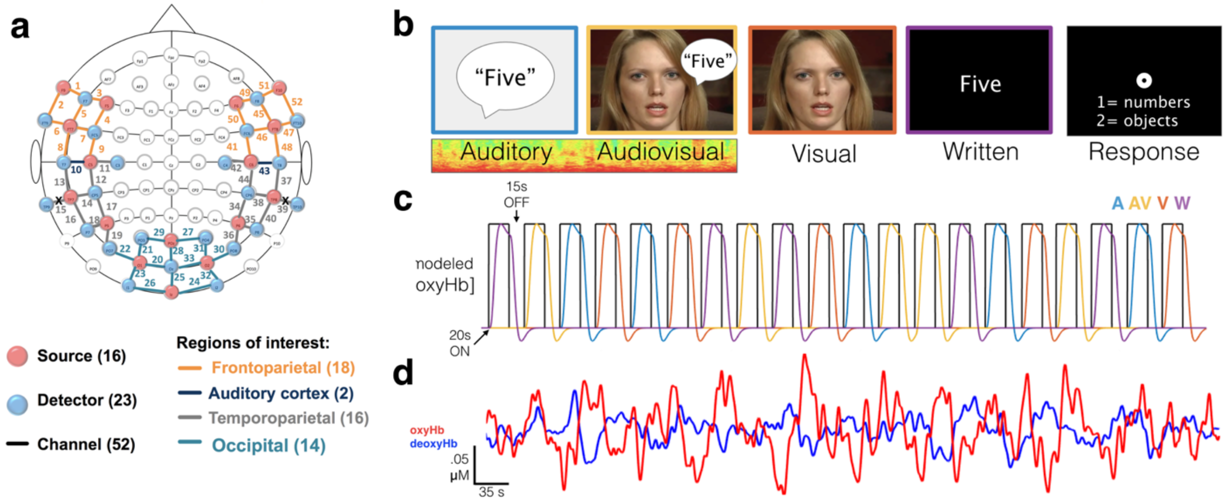


FIGURE 3.1. fNIRS task overview. Four regions of interest were defined across the 52 recording channels (d). Two channels (15 and 39) near the temporal bone were excluded from all analysis due to poor scalp contact. During each run, participants were presented with words in four conditions (b) from one of two categories (i.e., numbers v. objects or actions v. animals). Five words were presented in each 20 second block, after which the hemodynamic response function was allowed to return to baseline (c). Following preprocessing, an example time course for one (of the two 12 min) runs is shown in (d).

Subsequent data processing was done using Homer 2.0 software (Huppert et al., 2009) via Matlab 2017a (Mathworks). The Homer functions and their relevant parameters were: 1) enPruneChannels (0.5, 3.5, a.u.), 2) hmrIntensity2OD, 3) hmrMotionCorrectTDDR (Fishburn et al., 2019), 4) hmrBandpassfilt, (0.2, 0.01Hz), 5) hmrOD2Conc (default partial pathlength factor of 6) hmrBlockAvg (-2s to 35s). Together, these steps: eliminated noisy channels, converted raw intensities to optical density values, corrected for motion artifacts, filtered out cardiac and pulmonary artifacts, calculated concentration changes with a modified Beer Lambert law, and averaged all of these changes over the stimulus presentation blocks. The resulting signal changes (Fig. 3.1d) for each channel were analyzed with a general linear model (glm) using the Matlab function regress. The modeled signal described above and shown in Fig. 3.1c was entered into this regression. Four resulting beta (β) weights (one per condition) were analyzed across the group for significant activity using one tailed, one sample t-tests. Corrections for multiple comparisons are made using false discovery rate (FDR) within each ROI.

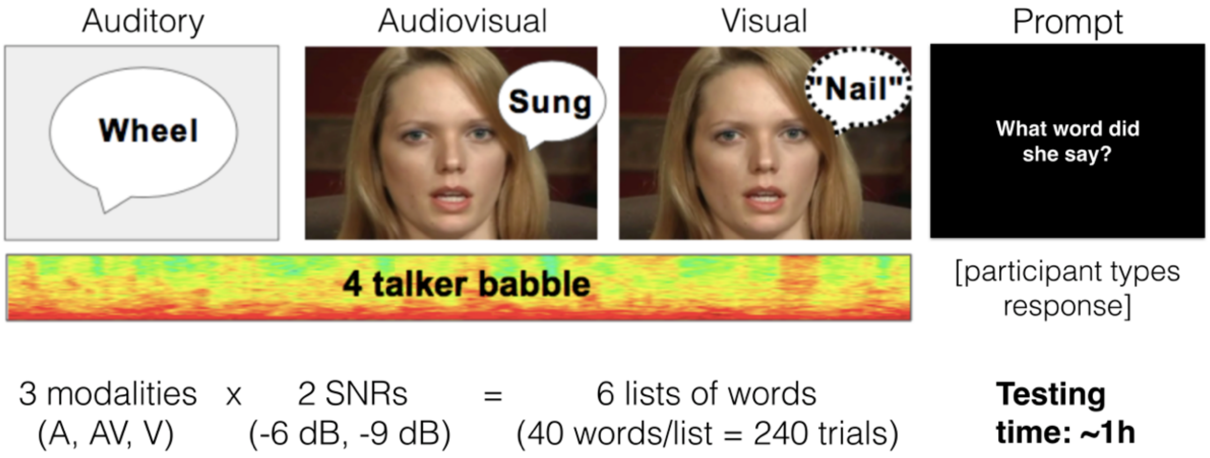
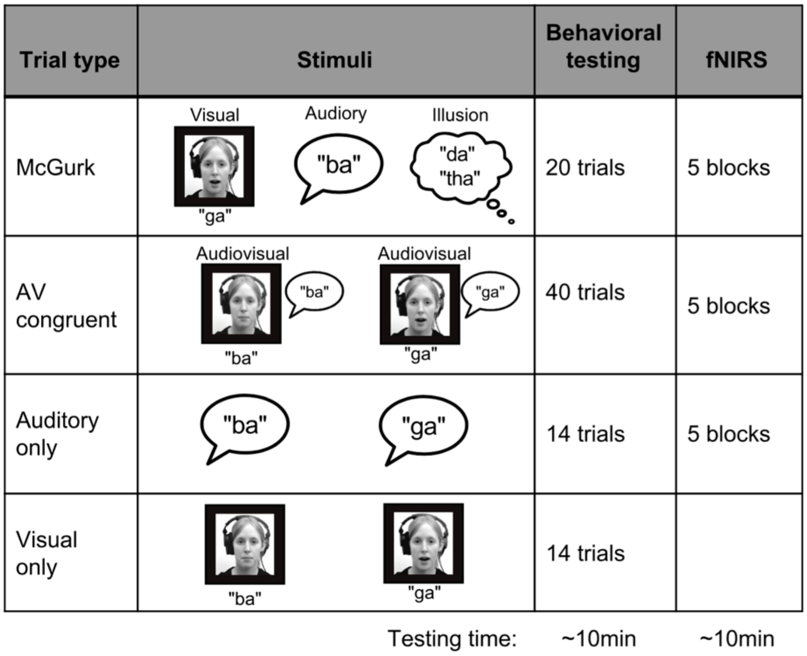


FIGURE 3.2. Word recognition task overview. For a behavioral correlate of audiovisual speech integration, participants were tested for monosyllabic word recognition in three modalities and two SNRs.

Word recognition in noise. For a behavioral measure of integrative speech ability, we designed a word recognition task with background noise at two signal-to-noise ratios (SNRs). Participants listened for a 60 dB monosyllabic target word centered within 2.5s of 4 talker babble at either 66 or 69 dB HL (i.e., -6 and -9 dB SNRs). These stimuli were created by Picou and colleagues who also did intelligibility balancing to ensure that words lists were well matched for listening difficulty (2011). In total, we tested 6 lists of 40 words each (Fig. 3.2). We quantified the percent of words correctly identified in each condition, and interactive index (ii) was calculated as a measure of audiovisual gain using the formula: $ii = [\ AV - \max_{(A,V)}\ ] / AV \times 100\%$. Results were analyzed using resampling based on a Welch's t-test in order to account for different sample distributions.

McGurk effect. We also compared behavioral and neuroimaging measures of AV speech integration to perception of the McGurk illusion in this cohort. As discussed previously, we measured perception of unisensory controls conditions (i.e., auditory only and visual only syllables) as well as congruent AV trials (Fig. 3.3). Participants responded to the question, "What did you hear?" using a keypad with the 4 options: /ba/, /ga/, /da/, or /tha/. Probably of perceiving the illusion was defined as both /da/ and /tha/ responses, henceforth simply referred to as /da/. We calculated the probability of perceiving the illusion [p(McGurk)] using a formula that subtracts incorrect /da/ responses in unisensory trials from /da/ responses to AV McGurk trials: p(McGurk| "da") x [1 – p(Unisensory|"da")].

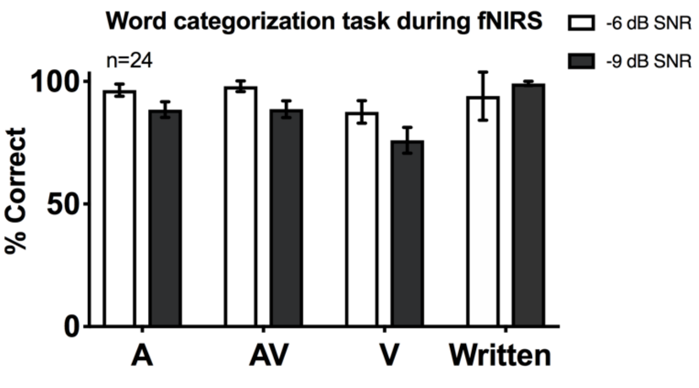


Trial type	Stimuli	Behavioral testing	fNIRS
McGurk	Visual "ga" Audiory "ba" Illusion "da" "tha"	20 trials	5 blocks
AV congruent	Audiovisual "ba" "ba" Audiovisual "ga" "ga"	40 trials	5 blocks
Auditory only	"ba" "ga"	14 trials	5 blocks
Visual only	"ba" "ga"	14 trials	
	Testing time:	~10min	~10min

FIGURE 3.3. McGurk experiment overview. Four trial types in behavioral experiments tested unisensory and audiovisual syllable identification as well as the illusory perception of the novel syllables /da/ or /tha/. One fNIRS run measured cortical activity during passive listening to the auditory and AV conditions.

For the fNIRS McGurk task, participants passively listened to three different conditions: McGurk trials, congruent AV stimuli, and unisensory syllables (Fig. 3.3). These data were analyzed in the same manner as the fNIRS word categorization task. Due to skewness in several measures, correlations between neuroimaging and behavior were assessed via Kendall's tau

3.3 Results



fNIRS. During fNIRS recordings, participants scored well above chance, suggesting that they were actively attending to stimuli throughout the experiment (Fig. 3.4). Of the 50 recording channels included in the analysis (Fig. 3.1a), there was significant activity for 11 (see Table 3.1).

FIGURE 3.4. Results for the word categorization task during fNIRS. This task was designed simply to maintain attention during fNIRS, and results indicate high accuracy across all conditions for each word group and signal-to-noise ratio (i.e., numbers v. objects for the -6 dB SNR and action verbs v. animals for -9 dB). Chance performance is at 50%,

Ch	MNI [x y z]	BA	Anatomical area	Condition	t statistic	p value
10	-48 -14 9	L 41/42	Primary auditory	Auditory Audiovisual	$t_{(22)} = 2.98$ $t_{(22)} = 3.87$	$p = 0.0034$ $p = 0.0004$
43	56 -10 7	R 41/42	Primary auditory	Auditory Audiovisual	$t_{(22)} = 2.42$ $t_{(22)} = 3.10$	$p = 0.012$ $p = 0.0026$
48	70 0 -13	R 22	Middle temporal gyrus	Auditory Audiovisual	$t_{(22)} = 3.69$ $t_{(22)} = 3.23$	$p_{corr} = 0.011$ $p_{corr} = 0.012$
4	-33 24 20	L 44/45	Broca's area	Auditory	$t_{(21)} = 2.75$	$p_{corr} = 0.040$
6	-51 4 -23	L 45	Broca's area	Audiovisual	$t_{(23)} = 4.16$	$p_{corr} = 0.003$
8	-51 -7 -9	L 22	Superior temporal	Auditory Audiovisual	$t_{(23)} = 2.69$ $t_{(23)} = 3.97$	$p_{corr} = 0.040$ $p_{corr} = 0.003$
27	21 -80 42	L 19	Visual association	Visual Written	$t_{(20)} = 5.44$ $t_{(20)} = 3.48$	$p_{corr} = 0.0002$ $p_{corr} = 0.017$
29	-15 -78 39	R 19	Visual association	Visual Written	$t_{(23)} = 3.12$ $t_{(23)} = 2.63$	$p_{corr} = 0.017$ $p_{corr} = 0.053$
17	-52 -51 29	L 39	Angular gyrus	Visual	$t_{(22)} = 2.78$	$p_{corr} = 0.049$
34	60 -49 31	R 39	Angular gyrus	Visual	$t_{(23)} = 2.68$	$p_{corr} = 0.049$
36	46 -68 15	R 19	Area MT	Visual	$t_{(20)} = 2.56$	$p_{corr} = 0.049$

TABLE 3.1. Significant fNIRS channels. In total, 11 channels had significant activity in the group. Ch = channel, BA = Brodmann area, MT = middle temporal.

Five channels (including the left and right primary auditory cortices) were significant during the auditory-only listening condition (Fig. 3.5). Of these channels, the four in the temporal lobe were also significant during the AV condition (Fig. 3.6). In addition, there was higher activity in channel 6 in the left hemisphere during AV listening.

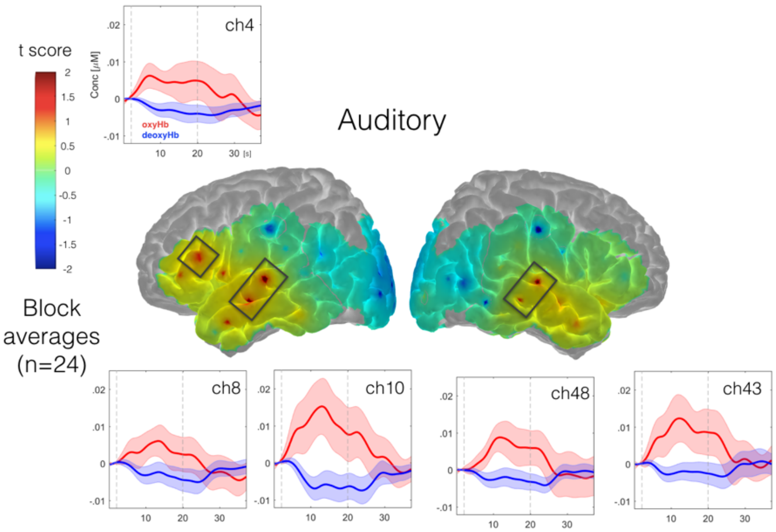


FIGURE 3.5. Auditory-evoked cortical activity. T scores are plotted as anatomical projections. Boxes indicate areas of significant activity and oxyHb (red) and deoxyHb (blue) concentration changes are shown for the 5 significant channels in these areas. Shaded regions correspond to 95% confidence intervals for the group.

Similar to the auditory condition, block averages corresponding to significant channels in the glm displayed a typical increase in oxyHb concentration that returned to baseline shortly after the stimulus ended (vertical dashed lines in Fig. 3.6 insets). In contrast, deoxygenated hemoglobin concentrations have a lower magnitude change in the opposite direction. In both the auditory and AV conditions, the left primary auditory channel (ch 10) had the largest relative increase in oxyHb concentration.

In the visual and written conditions, activity was primarily localized to the occipital cortex (Fig. 3.7). Lipreading evoked activity in both the visual cortex (ch 27 and 29) and middle temporal areas typically associated with visual motion perception (ch 17, 34, and 36).

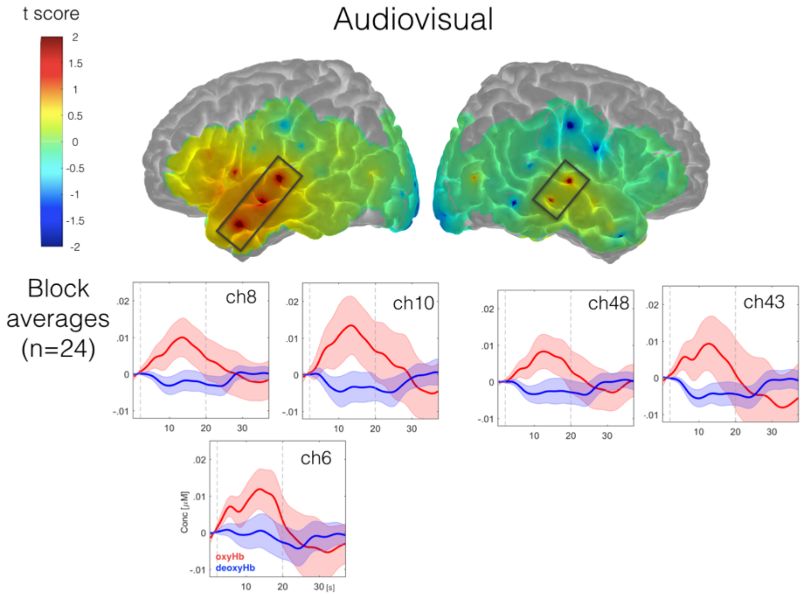


FIGURE 3.6. Audiovisual-evoked cortical activity. 5 channels bilaterally had significant AV activity, most of which were also identified in the auditory condition. Red = [oxyHb], Blue = [deoxyHb].

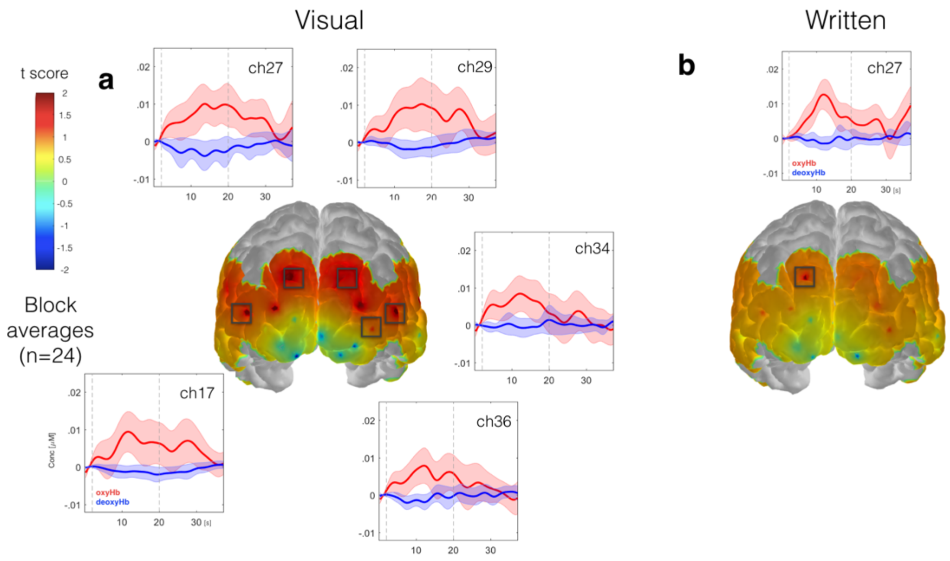


FIGURE 3.7. Cortical activity associated with visual (lipreading) and written word presentations. Red = [oxyHb], Blue = [deoxyHb].

Word recognition. In the behavioral word recognition task, lipreading ability was similar between the two noise levels (Fig. 3.8a). Group means are 14% for the 66dB noise level and 16% for the 69dB noise level (i.e., out of a possible 100%). During auditory-only listening, mean performance is 24% in the easier SNR, which is 21 percentage points lower than the 45% correct in the more difficult SNR condition. Simply adding the performance in these auditory and visual conditions would result in 58% and 41% words correct; however, the superadditive nature of multisensory integration results in the higher AV performance of 87% and 65% for the -6 and -9 dB SNRs, respectively. In comparing AV integration between the two conditions, ii was significantly higher in the louder noise level ($t_{(25.4)}$ = -2.3 , p_{adj} = 0.005), indicating greater benefit.

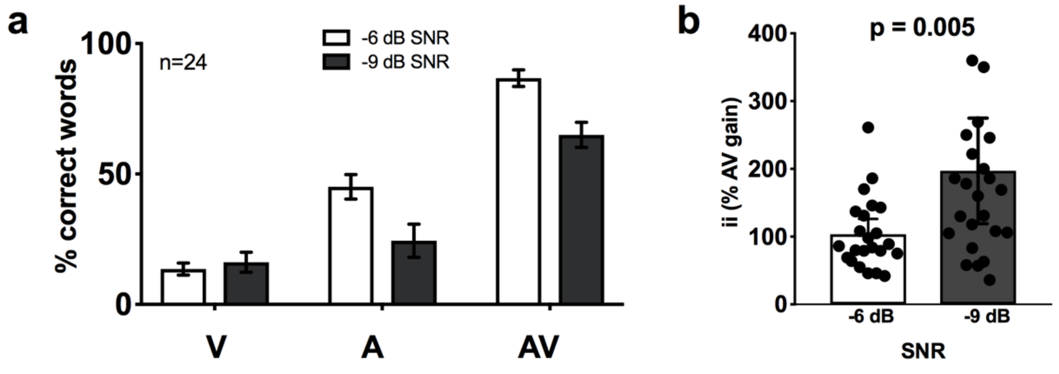


FIGURE 3.8. Audiovisual word recognition results. Error bars indicate 95% confidence intervals of the mean for the words identified in each list (a) and the interactive index for each SNR (b). Note that for better visualization one ii data point at 960% in the -9 dB SNR condition is omitted from panel (b).

McGurk tasks. In the control trials, there was nearly identical, ceiling performance for auditory-only and audiovisual identification of /ba/ and /ga/ (Fig. 3.9a). However, /ga/ is more visually-ambiguous, and there is significantly lower lipreading performance than the bilabial articulation of /ba/ ($t_{(23)} = 4.46$, $p = 0.0002$). The McGurk trials elicited responses of each component stimulus and the novel syllables (Fig. 3.9b). The average probability of perceiving the illusion was 0.36 on a scale of 0 to 1 where 1 is all /da/ responses to McGurk stimuli and no /da/ responses to any unisensory stimuli (Fig. 3.9c).

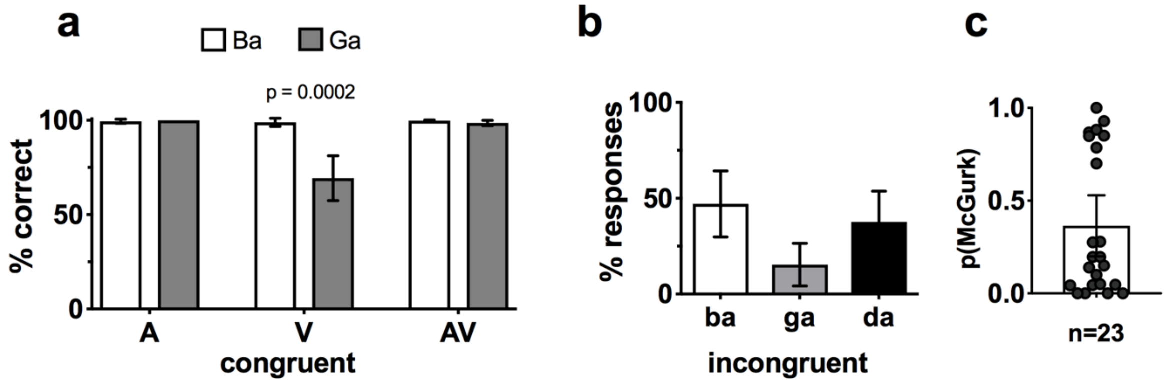


FIGURE 3.9. McGurk behavioral task results. Control trials of unisensory and congruent AV syllable identification were at or near ceiling with the exception of lipreading /ga/ (a). Responses to the McGurk trials were distributed among the auditory token /ba/, followed by the fused token, and lastly, the visual token /ga/ (b). The probability of perceiving the illusion has a large, somewhat bimodal distribution (c).

In the fNIRS task with passive listening to McGurk stimuli, there are two channels with significant activity (43 and 48), but only in the AV congruent condition (Fig. 3.10a). Block averages are plotted for all three conditions at these locations which are in the right hemisphere near auditory cortex (Fig. 3.10b).

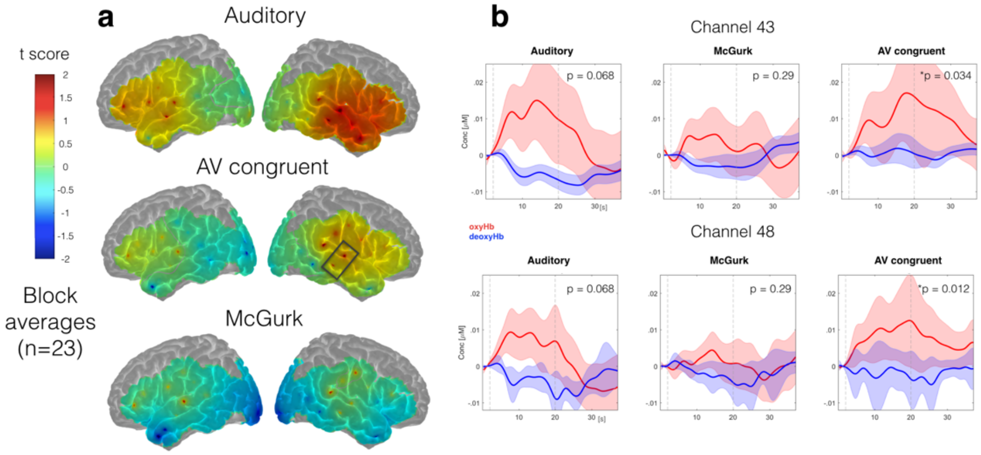


FIGURE 3.10. Cortical activity to McGurk stimuli. T scores are plotted for all three conditions (a). Across all channels, only two had significant activity for just the congruent AV syllables (b).

Correlations. p(McGurk) has a positive correlation to activity at channel 17 in the visual condition (tau = 0.45, p = 0.005). Additionally, ii in the -9 dB SNR condition had a negative correlation with cortical activity at channel 43 in the right auditory cortex during AV listening (tau = -0.30, p = 0.047). No other fNIRS channel correlated with behavioral measures of lipreading ability, p(McGurk), or ii.

3.4 Discussion

In this study, we accomplished our primary goal of designing behavioral and neuroimaging tasks to assay significant AV speech integration. That is, we measured higher AV gain (ii) as more background noise was added in a behavioral task, and we identified significant, left-lateralized cortical activity in both auditory-only and AV listening conditions in an fNIRS task. We saw the largest relative increase in oxyHb concentration for channel 10, which was nearest to the left primary auditory cortex. The difference in fNIRS-derived neural activity between A and AV listening conditions was, however, subtle. In future studies, it may be beneficial to add louder competing background noise in order to make this contrast more pronounced (Callan et al., 2003; Stevenson and James, 2009).

Written stimuli that required participants to silently read individual words only elicited activity in a primary visual area (Fig. 3.7b). Interestingly, silent lipreading recruited these and other visual areas likely related to motion processing (Fig. 3.7a). Given prior studies suggesting enhanced lipreading in CI users, it would be particularly interesting to compare activity at these locations to normal hearing controls.

We also sought to investigate the McGurk effect in the context of fNIRS imaging and the aforementioned word recognition task. We found no significant cortical activity in blocks of McGurk trials during fNIRS, but we did see right-lateralized AV activity in the congruent presentations of /ba/ and /ga/. This finding suggests different processing pathways between these two conditions; however, an active task would likely increase the overall percent signal change in these conditions to make this contrast a more robust comparison. Interestingly, p(McGurk) did have a positive correlation to activity at channel 17 in the visual condition. This may be functionally relevant for AV integration, considering the proximity to visual motion areas and multimodal integration in the area MT and posterior STS. Indeed, both auditory association areas and extrastriate visual areas provide input to regions of STS that are also implicated in illusion perception and, more broadly, AV speech integration (Lewis and Essen, 2000; Nath and Beauchamp, 2011, 2012; Seltzer et al., 1996). It would be interesting to see how this lipreading-evoked activity compares to CI users who others have suggested are better lipreaders but also less apt McGurk perceivers (Schorr et al., 2005).

Another interesting correlation in this study is a negative relationship between cortical activity at the right auditory cortex and interactive index at the more difficult SNR. It's possible that as noise increases, AV integration in the right hemisphere becomes more distributed and, consequently, more localized activity appears to be reduced. This relationship requires further characterization, particularly involving methods from which causality can be inferred. For instance, transcranial magnetic stimulation (TMS) in the vicinity of left STS is known to disrupt perception of the McGurk illusion (Beauchamp et al., 2010). It would be interesting, for instance, to test the effect of TMS application to the right STS on measures of AV gain. Finally, the absence of a correlation between McGurk results and word recognition results suggests that these two tasks may be more distinct than previously thought. Further work quantifying AV integration is more likely to have ecological relevance when assaying semantic speech tasks involving single words or full sentences.

3.5 Acknowledgments

Iliza Butera designed the experiments, collected data, analyzed the data, and wrote the manuscript. Andie DeFreese recruited participants and assisted in data collection. Baxter Rogers provided feedback on the experimental design. Mark Wallace and René Gifford provided guidance and feedback throughout the experiments and manuscript preparation.

CHAPTER 4 Behavioral and functional neuroimaging of CI users' audiovisual processing

4.1 Introduction

Although vision is known to play a critical role in communication, clinical assessments of CI candidacy and longitudinal postoperative outcomes have largely been limited to auditory-only measures. Consequently, current clinical assessments of CI outcomes are unable to describe the comprehensive profile of functional communication. We believe that a thorough investigation of visual abilities is the first step toward better understanding variability in ecological listening conditions that are inherently multisensory.

There is some evidence to suggest that multisensory integration in CI users may have greater perceptual benefits than what is experienced by normal-hearing individuals. In particular, a large longitudinal study of nearly 100 CI users found that visual speech proficiency was maintained over time and CI users outperformed normal hearing (NH) peers in their word-level AV integrative capabilities (Rouger et al., 2007). A caveat to this and several other studies directly comparing CI and NH groups is the use of noise vocoding to simulate the sound of an implant for normal hearing populations (Desai et al., 2008a). This technique may underestimate the unisensory identification of vocoded speech, potentially biasing between-group comparisons of the integration of AV integration (Schwartz, 2010). In this study, we equate CI and NH groups by adding differing levels of multitalker background noise. In doing so, we can simulate a familiar, yet challenging listening condition to ask whether CI users and NH controls experience the same visual benefit of concurrent lipreading.

Additionally, despite numerous, consistent reports of CI visual biasing in a McGurk task. No study to date has related this AV measure of illusory syllable perception to semantic, word-level audiovisual gain that may be evoked conversationally. One recent study, however, has investigated the role of crossmodal plasticity in facilitating the McGurk effect (Stropahl and Debener, 2017). In this study, the authors report that crossmodal activation of CI auditory cortex in response to faces had a positive relationship with the degree of McGurk illusion perception. Interestingly, a similar effect was also seen with a moderately hearing impaired group, likely suggesting an early onset of these adaptive changes in CI users.

There are two primary aims of the present study: 1) to test whether perceiving the McGurk illusion relates to other measures of AV integration and 2) to test whether behavioral and cortical processes of visual speech perception differ between CI users and NH controls in a way that may be beneficial for AV integration. Our primary hypothesis is that adult CI users will display enhanced audiovisual gain compared to normal-hearing controls due to differences in visual function and the networks supporting cross-modal plasticity. That is, CI users with greater visual acuity for speech will also exhibit greater functional connectivity between auditory, visual and multisensory cortices that supports more efficient audiovisual integration in ecological listening conditions that are naturally multisensory.

Though the McGurk effect and AV integration are sometimes referred to synonymously, recent behavioral and neuroimaging studies have cast doubt on whether the incongruent and, therefore unnatural McGurk stimuli engage different brain networks (Van Engen et al., 2017). Consequently, measures of semantic AV gain may or may not correspond to results of a McGurk task, as has been the assumption for some time. Either way, a greater knowledge of the ways in which visual cues can facilitate auditory comprehension, and how these interactions may differ from patient to patient, will likely have important implications for both predicting CI outcomes and optimizing remediation of speech comprehension.

4.2 Methods

Participants. This study included 23 postlingually-deafened cochlear implant users and 23 age-matched, normal hearing controls (Table 4.1). All individuals in the normal hearing (NH) group passed hearing screenings to detect pure tones at age-appropriate levels for the following frequencies: 250, 500, 1000, 2000, and 4000 Hz. For CI users, we calculated a "duration of deafness," which we defined as the date of cochlear implantation minus the prior onset of severe hearing loss. For a practical estimate of the onset of severe hearing loss, we asked participants if they remembered approximately how old they were when they could no longer talk on the phone without visual aids. The average duration of deafness was (8.1y, SD= 10.1).

All participants self-reported normal or corrected-to-normal vision, no major psychological or neurological disorders, and groups were matched for handedness (1 left-hand dominant, 22 right-hand dominant per group). There were no between-group differences in age ($t_{(44)} = -0.51$, $p = 0.96$; Fig. 1c) or nonverbal IQ ($t_{(41)} = 0.26$, $p = 0.95$) as measured by the Kaufman Brink Intelligence Test (see Table 4.1). All participants were native English speakers, and all testing took place at

Vanderbilt University Medical Center between May 2017 and October 2018. All participants provided informed consent prior to any testing and all procedures were approved by Vanderbilt University's Institutional Review Board.

Group	N	Sex (% female)	Mean age (y)	Number of CIs	Acoustic hearing	nonverbal IQ
CI	23	70%	50.9 ± 12.2	1 – 91% 2 – 9%	65%	108 ± 10
NH	23	73%	51.0 ± 12.0	—	100%	107 ± 11

TABLE 4.1. Participant characterization. Values are means ± standard deviation.

Auditory testing. Cochlear implant users completed all testing in their "best-aided" condition, which included hearing aid amplification for 15 individuals (i.e., 65% of CI users). Standard clinical speech testing of monosyllabic, consonant nucleus consonant (CNC) word lists was tested at 60 dB HL without noise for all CI users, and subsequent testing at +10 and +5 dB signal-to-noise ratios (SNRs) was also completed with a subset of high performers (Fig. 4.1a). More extensive auditory testing was done in the normal hearing control group in order to derive full psychometric curves from which any performance level from their CI counterparts could be estimated (Fig. 4.1b). This testing included 260 randomized trials of 20 trials per 1 dB increments for 13 SNRs (i.e., from -10 dB to +2 dB), lasting approximately 30 minutes.

Word recognition in noise. To quantify audiovisual benefit, we tested word recognition performance in the presence of multi-talker background noise at various sound levels. Noise stimuli consisted of 2.5s clips of 4 female speakers concurrently reading different children's encyclopedia entries (Picou et al., 2011). The target word within this noise was monosyllabic, 1.7s in duration, and spoken by a female adult (Fig. 4.2a) (Picou and Ricketts, 2014).

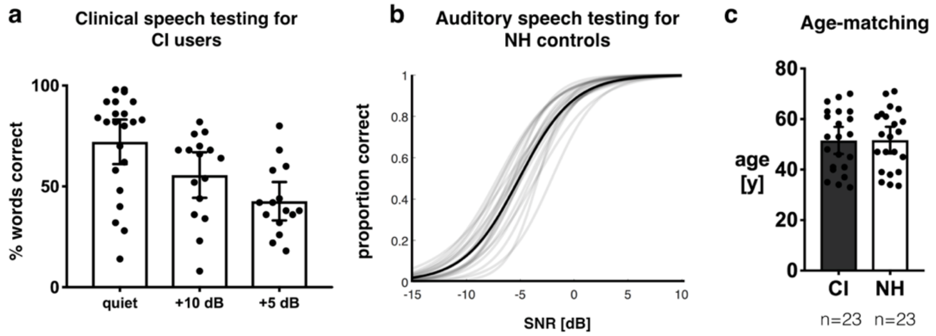
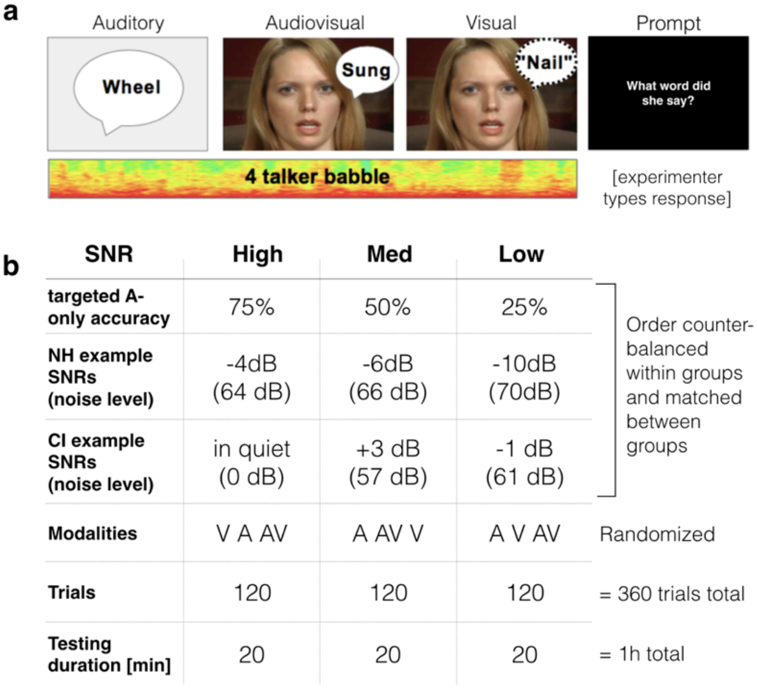


FIGURE 4.1. Auditory speech testing and age-matching between groups. Standard clinical testing of monosyllabic (CNC) words was used to characterize auditory-only listening performance in CI users (a). In normal hearing (NH) controls, 13 SNRs were tested using experimental stimuli to fit psychometric functions for each individual (gray). These individual curves were used to select noise levels to equate performance with 23 age-matched CI users (c).

In total, we tested 9 lists (3 SNRs x 3 modalities), each consisting of 40 words that were balanced for intelligibility (Picou et al., 2017). (See Appendix B for all words and Appendix C for all list balancing estimates). 286 words were presented across three modalities (auditory-only wav files, visual-only lipreading avi files, and audiovisual avi files), which were pseudo randomized within each SNR using E prime 2.0 software (Schneider et al., 2002). After each trial, participants repeated the word aloud, and the experimenter typed it on the screen to confirm accuracy. Sixty-two of the words (17%) were repeated once for a total 360 trials, which took approximately 1h to complete (Fig. 4.2b).

After creating a long-term average spectrum (LTASS) from all the words (Donley et al., 2018; Donley, Jacob, 2017), we used a Larson Davis sound level meter to calibrate the target words to 60dB HL. The sound level of the "babble" was calculated in a similar manner in order to create noise files corresponding to SNRs between -12 dB and +15 dB. Lastly, we quantified the percent of full words correctly identified in each condition, and interactive index (ii) was calculated as a measure of audiovisual gain using the formula: $ii = [\ AV - \max_{(A,V)}\ ] / AV \times 100\%$. Because ii is calculated from performance in the AV condition relative to whichever unisensory condition is highest, we questioned whether measuring AV gain strictly relative to auditory performance—a metric sometimes referred to as visuoauditory benefit (Rouger et al., 2007) or visual enhancement (Grant and Seitz, 1998, 2000b; Sommers et al., 2005; Van Engen et al., 2017)—might differ between groups. Therefore, we also calculated gain using the formula $(AV - A) / (100 - A)$.



SNR	High	Med	Low	
targeted A-only accuracy	75%	50%	25%	Order counter-balanced within groups and matched between groups
NH example SNRs (noise level)	-4dB (64 dB)	-6dB (66 dB)	-10dB (70dB)	
CI example SNRs (noise level)	in quiet (0 dB)	+3 dB (57 dB)	-1 dB (61 dB)	
Modalities	V A AV	A AV V	A V AV	Randomized
Trials	120	120	120	= 360 trials total
Testing duration [min]	20	20	20	= 1h total

FIGURE 4.2. Experimental design for word recognition testing. Monosyllabic words spoken by an adult female were presented in three modalities (a). Additionally, three signal-to-noise ratios (SNRs) were selected for each individual and counterbalanced between groups (b). Modalities were pseudo randomized within each SNR for a total of 360 trials and about an hour of testing

McGurk Effect. We tested visual-only, auditory-only, and congruent audiovisual presentations of the syllables 'ba' and 'ga' using videos previously demonstrated to be effective in eliciting the illusion (Quinto et al., 2010). The experimental design, described in prior work from our lab (Stevenson et al., 2012), includes the presentation of unisensory control conditions in a separate block prior to audiovisual trials. McGurk stimuli consisted of incongruent pairings of the phoneme 'ba' dubbed over the viseme 'ga.' These stimuli were pseudorandomized among congruent AV trials. After each trial, participants responded to the question, "What did you hear?" using a keypad with the 4 options: 'ba', 'ga', 'da', or 'tha'. Testing lasted approximately 10 minutes.

In the unisensory block, responses were averaged over 7 trials per phoneme/modality combination for 28 trials total. In the AV block, 20 trials were averaged per condition (McGurk stimuli, and congruent 'ba' and 'ga') for 60 total trials. We corrected for erroneous reports of the fused syllables ('da' or 'tha') in unisensory conditions—a likely issue for CI users and one that could overestimate the magnitude of the perceived illusion in McGurk trials. This correction involved calculating the probability of 'da' reports in the McGurk trials relative to a unisensory baseline: $p(McGurk|da) \times [1 - p(Unisensory|da)]$. This calculation ensures that estimates of the McGurk effect are not merely due to incorrect identification of unisensory components as the fused syllables.

Audiovisual McGurk data from 4 CI users were excluded due to a computer timing issue caused by a failing power supply. A delay in the visual content is likely to have impacted results for these individuals on account of the reduced prevalence of the McGurk effect with AV asynchrony above 300ms (van Wassenhove et al., 2007).

fNIRS experimental design. For neuroimaging, sound and video files were selected from the same stimulus bank as the word recognition task (Picou et al., 2011). We tested cortical activity in response to 4 types of stimulus presentations: auditory-only listening in noise, audiovisual listening in noise, silent visual-only lipreading, and silent reading (Fig. 4.3a). We presented five words per block after which participants had 2.2s to respond with the numbers 1 or 2 on a keypad in order to identify the category of each word (see Appendix A for a list of all words). The categories were numbers v. objects in Run 1 (e.g., bell = 1, eleven = 2) or animals v. actions in Run 2 (e.g., bird = 1, walk = 2). A list of all the words was displayed at the beginning of each experiment and participants confirmed understanding after a practice trial. After each 20s block, participants were instructed to maintain stillness and fixation for 10s, which served as a baseline period before the next stimulus. The four modalities (A, V, AV, written) were repeated 5 times per run for a total of 20 blocks and ~10 min of imaging time each. The primary purpose of the categorization task was to keep participants actively attending to stimuli for the duration of the session, which took place in a dark and quiet room at the Vanderbilt Institute of Imaging Sciences (VUIIS).

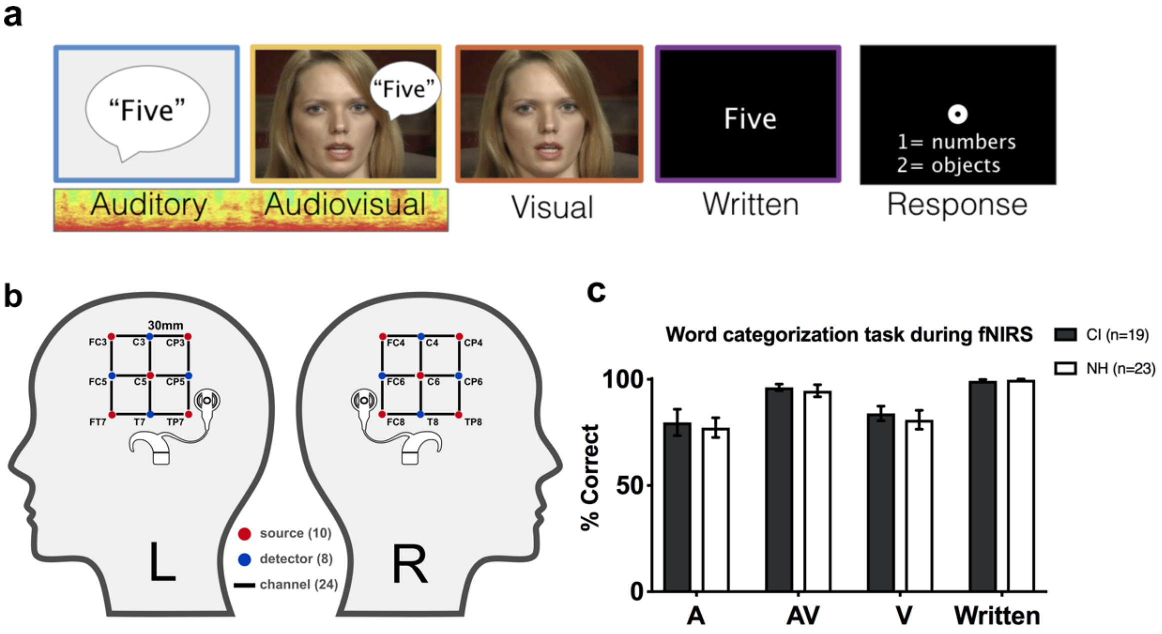


FIGURE 4.3. Functional neuroimaging of audiovisual speech. Participants categorized words that were presented in four modalities (a) during bilateral, 24 channel fNIRS recordings using a Hitachi ETC-4000 (b). Behavioral results (c) confirm that participants were attending to stimuli in each condition (i.e., results are above chance at 50%).

For neuroimaging, we used a continuous wave Hitachi ETG-4000 for bilateral fNIRS recordings at a sample rate of 10 Hz (Fig. 4.3b). 10 infrared light sources emitting 695 and 830 nm wavelengths were centered over the 10-20 points C5 and C6 on the left and right hemispheres, respectively. Repeatable probe placements were done by taking measurements for each individual from nasion to and between the two prearicular points. From these measurements, we identified Cz and the center of each 3x3 probe holder as 20% from the prearicular as C5/6. We chose this placement in order to record from auditory and audiovisual speech processing areas of the temporal lobe while avoiding cochlear implant receiver coils (Fig. 4.3b). Because there is variability in the receiver locations, neighboring channels were occasionally affected, which were pruned from the analysis (see below).

In addition to the 10 light sources, 8 detectors were spaced 30mm apart to create 24 recording channels. Stimuli were presented with E prime 2.0 software and concurrent triggers were sent via serial port to the recording software in order to align optical data with stimulus events.

fNIRS analysis. Data were analyzed using Homer 2.0 (Huppert et al., 2009) via Matlab 2017a. Infrared light travels well through skin and bone but not though hair, particularly darker hair colors that contain higher concentrations of light-absorbing pigments. As a result, care was taken to part hair around channels whenever possible. Because excessively noisy channels did occur on account of poor scalp contact from hair or occasionally implants, channel pruning was done using the function enPruneChannels and the conservative cutoff reflectance values of 0.5 to 3.5. These criteria were ~4 SD from mean values, which resulted in pruning just 13 channels across all recording sites in the NH group (i.e., 1.2%) and 24 channels (5%) in the CI group.

Following initial channel pruning, intensity values were converted to optical density using the function hmrIntensity2OD. Next, we corrected for motion artifacts using a regression-based Temporal Derivative Distribution Repair or TDDR function hmrMotionCorrectTDDR (Fishburn et al., 2019). Unlike other motion correction approaches, TDDR makes minimal assumptions about the data while correcting scalp decoupling from optodes as a result of motion, which appears as spikes or step-like shifts in the baseline (see Appendix D for examples and TDDR corrections). Additionally, systemic artifacts like cardiac and respiratory oscillations were removed by bandpass filtering from 0.01 to 0.2 Hz. The block design of the experiment has a frequency of 0.033 Hz (30s period), which is sufficiently within this filter range.

The final steps of fNIRS analysis involved converting optical density measures to concentrations via a modified Beer-Lambert law and the function hmrOD2Conc. We used the default partial pathlength factor of 6mm for this conversion, and subsequent block averaging was done from 2s prior to the stimulus onset to 9s after via the hmrBlockAvg function.

Procedures. Auditory testing was completed in an isolated sound booth with words presented via a monospeaker positioned directly in front of participants (i.e., 0° azimuth). Sound levels were calibrated prior to each session and adjusted as needed using a standard audiometer. All participants were compensated with a $50 gift card for 4 hours of participation over the course of 1-2 days.

Statistical approach. In the word recognition task, the percent of words correctly identified in all modalities was compared between groups using independent samples t-tests with FDR corrections for multiple comparisons (i.e., 3 per modality). Adjusted p values are reported for significant group differences. To test for group differences with McGurk control trials, we used a mixed-model ANOVA (i.e., 3 modalities × 2 syllables × 2 groups) and independent samples t-tests for pairwise follow up of significant interactions. Comparing responses to McGurk trials, we needed a test with minimal assumptions of the distribution between groups, so we used a resampling approach in R. This script used the p value from a Welch's two sample t-test as the metric of comparison because it captures both the central tendency and variability.

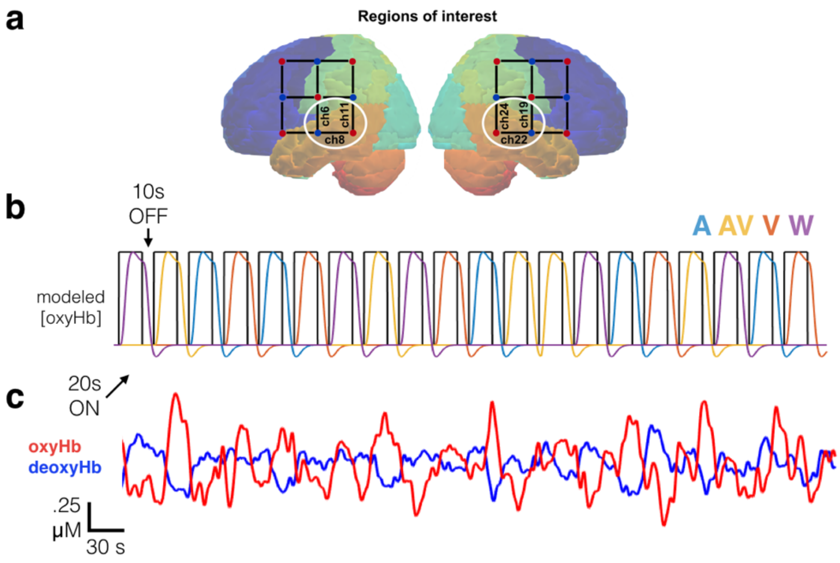


FIGURE 4.4. fNIRS analytical approach. Two regions of interest (ROIs) spanning 6 channels (a) were analyzed for an increase in the concentration of oxygenated hemoglobin along a timescale for a standard hemodynamic response function (b). An example time course for one individual depicts micromolar concentration changes for data recorded at channel 8, centered over primary auditory cortex.

fNIRS. First, we defined a region of interest nearest to speech processing areas of the temporal lobe that spanned 6 channels bilaterally (Fig. 4.4a). To test for significant cortical activity, we convolved the boxcar function of stimulus blocks with a standard hemodynamic response function (hrf) to create a model of the change in oxygenated hemoglobin (Fig. 4.4b). Next, we ran a regression between the model and the preprocessed data for each channel (Fig. 4.4c) using the built-in Matlab function regress. We averaged the resulting β weights between the two runs, totaling four β weights per channel per individual. Because the magnitude of these values is proportional to the amplitude of concentration changes, one-sample t-tests indicate significant activity for each group. FDR corrections were applied within each ROI and adjusted values are reported. For those channels meeting significance at $\alpha < 0.05$, block averages of the deoxyHb (blue) and oxyHb (red) concentrations are plotted with confidence intervals for the group.

Correlations. To evaluate the correlation between McGurk results, word recognition, fNIRS, and clinical measures we report Kendall's tau values. This nonparametric test evaluates the association between rank orders in order to compare metrics with inherent skew such as AV gain. Values range from -1 to +1 to represent perfect negative or positive relationships and no linear relationship indicated by 0. Two-tailed p values are reported for each metric.

4.3 Results

Word recognition task. On average the SNRs in the NH group were 12 dB louder than those selected for CI users (Fig. 4.5a). As expected, the percent of words correctly identified in auditory-only testing (Fig. 4.5b) is statistically-equivalent between groups (Table 4.2). Pairwise comparisons within each modality indicate that CI lipreading performance in two (of the three) noise levels is significantly higher than NH controls (Table 4.2). Individual data for matched pairs (shown in corresponding colors in Fig 4.6) were used to calculate interactive index (ii) as a measure of audiovisual gain (Fig. 4.6 d,h) (Stevenson et al., 2014b). Despite being better lipreaders, there are no corresponding increases in AV word recognition or ii for the CI group (see Table 4.2).

SNR	Modality	Mean ± SD	t statistic	Significance
High	A	CI 61.7% ± 25.9% NH 60.8% ± 24.2%	$t_{(44)} = 0.11$	p = 0.91
Med		CI 42.2% ± 14.4% NH41.1% ± 17.1%	$t_{(38)} = 0.22$	p = 0.83
Low		CI 21.9% ± 11.1% NH 24.7% ± 14.0%	$t_{(42)} = -0.75$	p = 0.46
High	V	CI 18.8% ± 8.2% NH 13.4% ± 5.9%	$t_{(44)} = 2.56$	**$p_{adj} = 0.021$**
Med		CI 16.0% ± 6.3% NH 12.8% ± 5.5%	$t_{(38)} = 1.73$	p = 0.092
		CI 24.4% ± 9.8% NH 16.4% ± 7.3%	$t_{(42)} = 3.07$	**$p_{adj} = 0.012$**
High	AV	CI 84.3% ± 14.4% NH 84.7% ± 13.1%	$t_{(44)} = -0.096$	p = 0.92
Med		CI 69.7% ± 11.9% NH 66.1% ± 14.2	$t_{(38)} = 0.86$	p = 0.39
Low		CI 64.4% ± 12.4% NH 63.1% ± 16.1%	$t_{(42)} = 0.29$	p = 0.77
High	ii	CI 45.3% ± 45.3% NH 80.7% ± 130.1%	$t_{(44)} = -1.23$	p = 0.23

Med	CI 79.7% ± 54.8% NH 84.8% ± 83.6%	$t_{(38)} = -0.23$	p = 0.82
Low	CI 157.0% ± 127.2% NH 175.5% ± 128.5%	$t_{(42)} = -0.48$	p = 0.64

TABLE 4.2. Between-group differences in audiovisual word recognition. The CI group scored significantly higher with visual-only lipreading in the high and low noise conditions (bold) after applying FDR corrections for multiple comparisons.

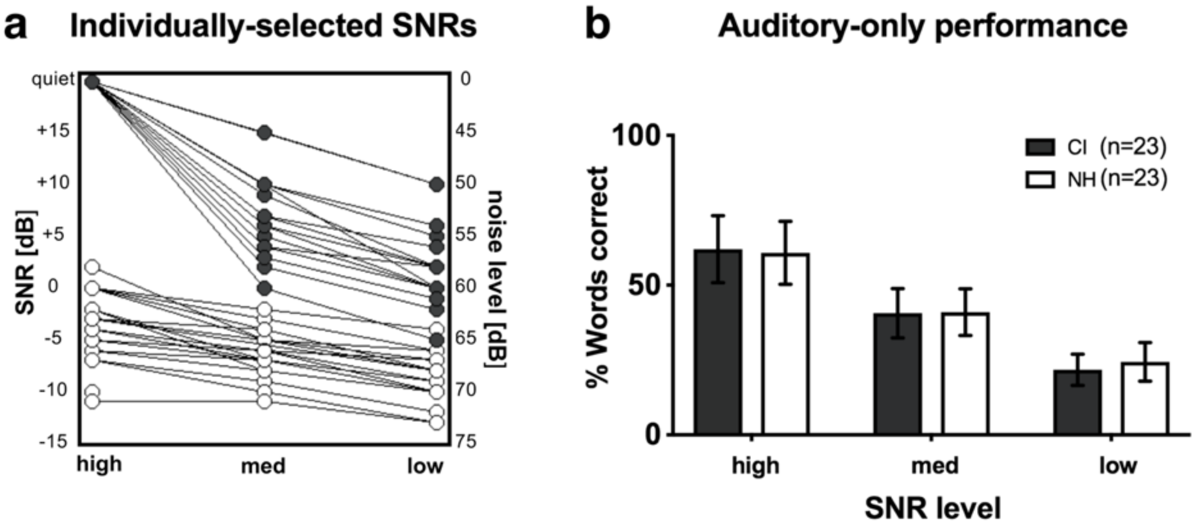


FIGURE 4.5. Adult CI users matched to NH controls. Signal-to-noise ratios (SNRs) were selected for each individual in order to equate groups for auditory-only performance. On average, NH controls (white) were tested in background noise that was 12 dB louder than CI user testing (gray) (a), which resulted in statistically-equivalent performance between the groups (b).

We found the magnitude of visual enhancement (see methods) to be slightly higher for CI users compared to ii; however, there are also no between-group differences (data not shown). This further suggests that CI and NH groups display a similar magnitude of integrative benefit of visual and auditory speech in background noise. It should be noted, however, that for the most difficult SNRs tested, the modality with the highest accuracy was auditory for 60% of NH controls and visual for 65% of CI users (Fig. 4.7). In summary, when matching auditory-only word recognition, CI users are slightly better lipreaders in background noise, and oftentimes, this lipreading ability outperforms their auditory-only listening ability, resulting in similar AV performance and measures of multisensory gain.

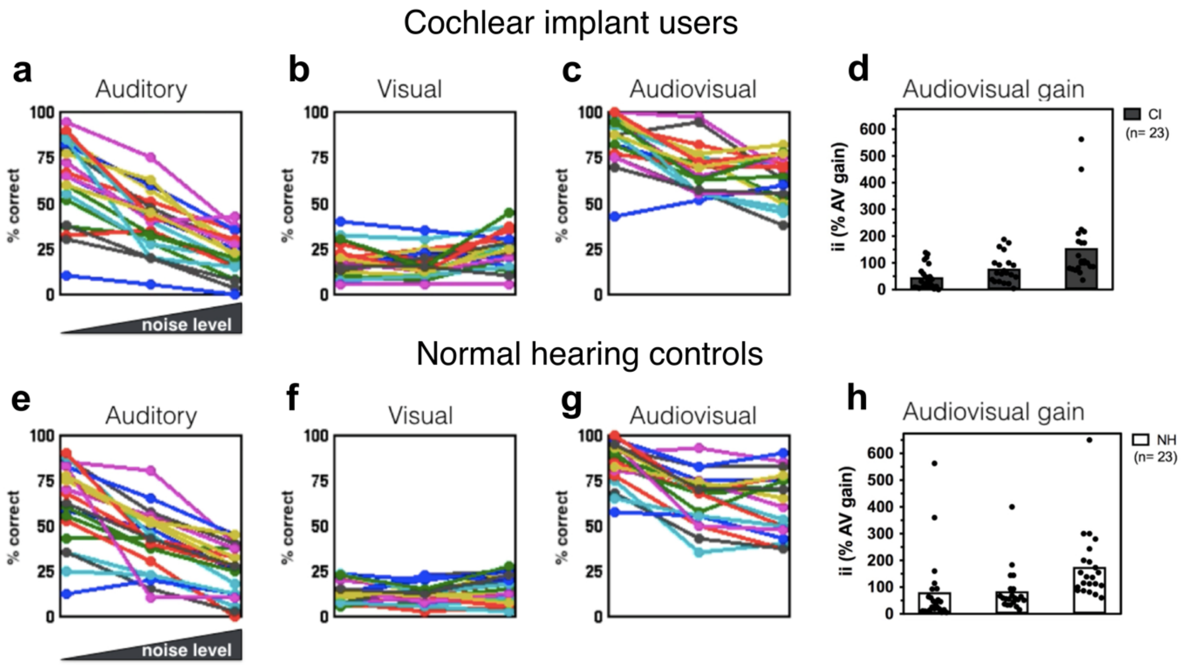


FIGURE 4.6. Word recognition results. The percent of words correctly identified is plotted for CI users (top) and NH controls (bottom) with colors corresponding to matched individuals between the two groups. SNRs are plotted with increasing difficulty for auditory-only testing (a, e), visual-only lipreading (b,f), and audiovisual testing (c,g). Interactive index—a measure of percent audiovisual gain—is also plotted for all three SNRs (d,h). Though CI users are better lipreaders, there are no significant differences in AV gain.

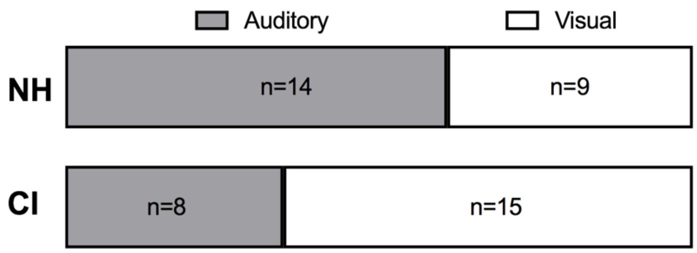


FIGURE 4.7. Modality with maximum unisensory performance. Although AV gain is not significantly different between the CI and NH groups, the modality with the maximum unisensory performance for the most difficult condition, is auditory for 60% of NH listeners and visual for 65% of CI users.

McGurk task. For controls trials in the McGurk task, a mixed-model ANOVA (i.e., 3 modalities × 2 syllables × 2 groups) indicates a very large effect of group ($F_{(1,40)} = 48.6$, $p < 0.001$, $\eta^2 = 0.55$) as well as a modality × syllable × group interaction ($F_{(2,40)} = 35.3$, $p < 0.001$, $\eta^2 = 0.47$). This interaction suggests that there are group differences for one or more modality/syllable combinations, which necessitates pairwise tests. As a result of these follow up t tests shown in Table 4.3, the CI group had significantly lower auditory-only identification of both ba and ga as well as the AV identification of the visually-ambiguous syllable ga (Fig. 4.8a).

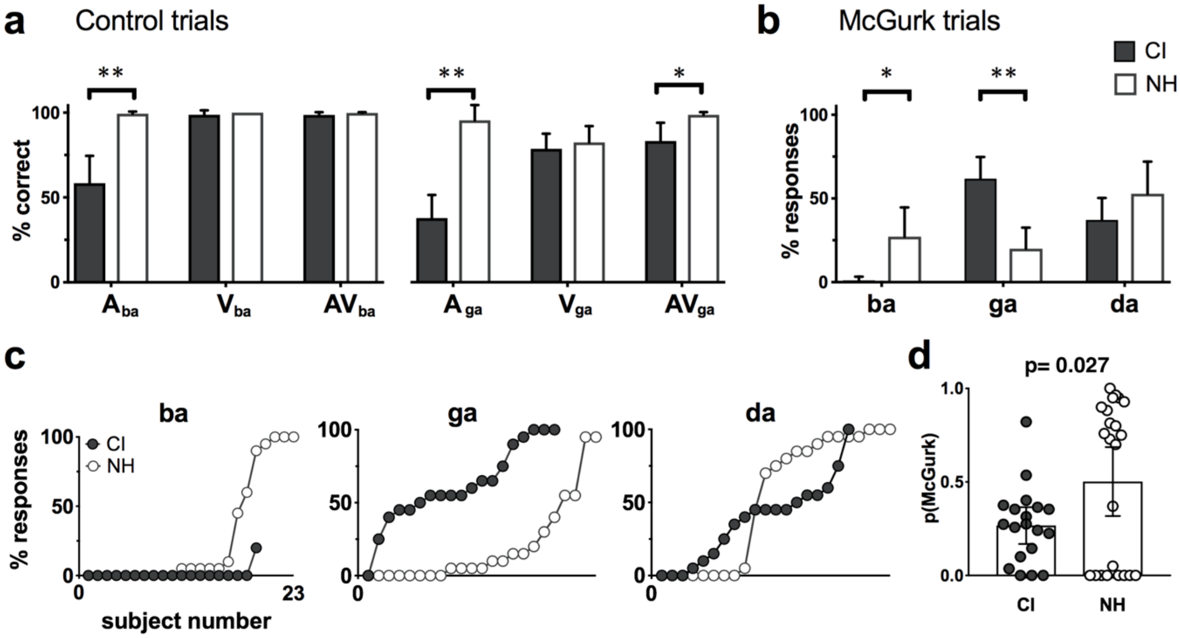


FIGURE 4.8. McGurk Illusion results. Control trials (a) tested accuracy for identifying the syllables 'ba' and 'ga' in unisensory and congruent audiovisual presentations. Mean responses to the incongruent McGurk trials (b) indicate CI user bias toward the visual component (ga), NH bias for the auditory component (ba), and yet similar reports of the illusory syllable (da/tha). Individual data for McGurk trials is shown in (c) and the derived probability of McGurk perception (d) both indicate a slightly lower, yet non-significant, difference in illusory reports CI group. Error bars indicate 95% confidence intervals of the mean. *p <0.01, **p< 0.001

For McGurk trials incongruently pairing the visual articulation of 'ga' with the audio of 'ba,' there was a syllable × group interaction ($F_{(2,39)} = 13.3$, $p < 0.001$, $\eta^2 = 0.41$). Pairwise follow up tests indicate: 1) there are significantly more reports of the syllable 'ba' in the NH group ($t_{(40)} = -2.79$, $p = 0.008$), 2) more reports of 'ga' in the CI group ($t_{(40)} = 4.81$, $p < 0.001$), and 3) no difference between groups for 'da' ($t_{(40)} = -1.34$, $p = 0.187$; Fig. 4.8b). Individual data from the McGurk trials are shown in Figure 4.8c and indicate a wider separation in the NH group between individuals who frequently perceive the illusion and those who do not. In order to correct for incorrect 'da' responses in the control conditions that could falsely inflate the magnitude of the illusion, the adjusted probability of McGurk perception is plotted in Fig. 4.8d. For NH controls, the mean p(McGurk) is the nearly same as 'da' reports (i.e., 0.528 [Fig. 4.7b] v. 0.502 [Fig. 4.8d], a 5% decrease). However, the CI group average for p(McGurk) is 0.267—a value 28% lower than their proportion of 'da' responses (mean = 0.371, Fig. 4.8b). A Welch's t test with resampling indicates significant group differences in p(McGurk) while making minimal assumptions of the underlying distributions between the two groups—a necessary precaution given the nearly bimodal distribution in the NH group (Fig. 4.8d). Results of this test are significant at $\alpha < 0.05$ both with and without resampling ($t_{(32.8)} = -2.34$, p =

0.025, p_{adj}= 0.027). Furthermore, repeating this test on strictly non-zero values (i.e., excluding the 3 CI users and 8 NH controls who did not perceive the illusion) indicates even more pronounced group differences of lower McGurk perception in CI users (mean = 0.317) compared to controls (mean = 0.770, $t_{(25.3)}$= -5.69, p = 6.1×10^{-6}).

Modality	Syllable	t statistic	Significance
A	ba	$t_{(44)} = -5.76$	**p = 7.7x 10^{-7}**
V		$t_{(44)} = -1.00$	p = 0.32
AV		$t_{(40)} = -1.53$	p = 0.14
A	ga	$t_{(44)} = -7.33$	**p = 3.8 x 10^{-9}**
V		$t_{(44)} = -0.59$	p = 0.56
AV		$t_{(40)} = -3.27$	**p = 0.002**

TABLE 4.3. Between-group differences in McGurk control conditions. The CI group score significantly lower with identifying auditory-only presentations of 'ba' and 'ga' as well as AV presentations of 'ga.' (bold)

In summary, CI users who don't perceive a fused McGurk syllable nearly exclusively report the visual component ('ga'); while non-fusing NH controls typically report the auditory component ('ba'; Fig. 4.8b). Though the two groups have similar responses of the fused percepts 'da' or 'tha', auditory ambiguity for CI users in controls conditions (Fig. 4.8a) calls for a corrective calculation of p(Mcgurk)—a metric that significantly differs between groups for all subjects (Fig. 8d), and particularly when excluding non-perceivers of the illusion in both groups. In both circumstances, CI users have a lower probably of perceiving the illusion.

fNIRS task. Behavioral results during fNIRS indicate that individuals in both groups actively attended to stimuli and correctly categorized words above chance (i.e., 50% accuracy; Fig. 4.3c). For the NH group, 4 recording channels had significant activity in three conditions: Auditory, Audiovisual, and Written (Table 4.4).

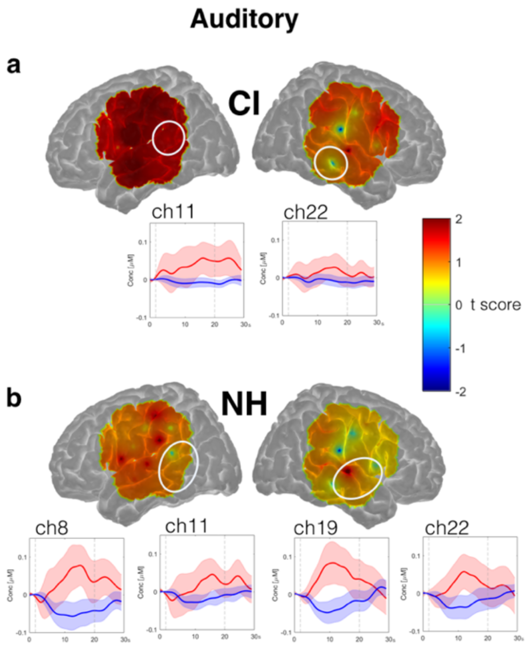


FIGURE 4.9. Cortical activity during auditory-only listening. The CI group (a) had two significant channels (white circles). Block averages indicate a modest magnitude of HbO (red) and HbR (blue) changes over time in units of micromolar. Solid lines = group mean, shading = 95% confidence intervals. NH controls (b) had broader elicited activity across 4 channels.

Group	Ch	MNI [x y z]	BA	Anatomical area	Condition	t statistic	p value
NH	8	-47 -28 -2	L 22	superior temporal	Auditory Written	$t_{(22)} = 2.63$ $t_{(22)} = 3.95$	$p_{corr} = 0.023$ $p_{corr} = 0.0041$
NH	11	-41 -36 16	L 40	supramarginal gyrus	Auditory	$t_{(22)} = 3.06$	$p_{corr} = 0.017$
NH	19	56 -10 7	R 41/42	primary auditory cortex	Auditory Audiovisual	$t_{(22)} = 3.47$ $t_{(22)} = 4.29$	$p_{corr} = 0.013$ $p_{corr} = 0.002$
NH	22	53 -26 -8	R 21	middle temporal	Auditory	$t_{(22)} = 2.85$	$p_{corr} = 0.019$
CI	11	-41 -36 16	L 40	supramarginal gyrus	Auditory	$t_{(21)} = 2.64$	$p_{corr} = 0.050$
CI	22	53 -26 -8	R 21	middle temporal	Auditory	$t_{(21)} = 2.89$	$p_{corr} = 0.050$

TABLE 4.4. Significant activity measured in 4 fNIRS channels. Ch = channel; BA = Brodmann area

For CI users, significant activity was only identified in the auditory-only listening-in-noise condition in one channel in the left and right hemispheres (Fig. 4.9a). These and 2 other channels were identified in the NH group for this listening condition (Fig. 4.9b).

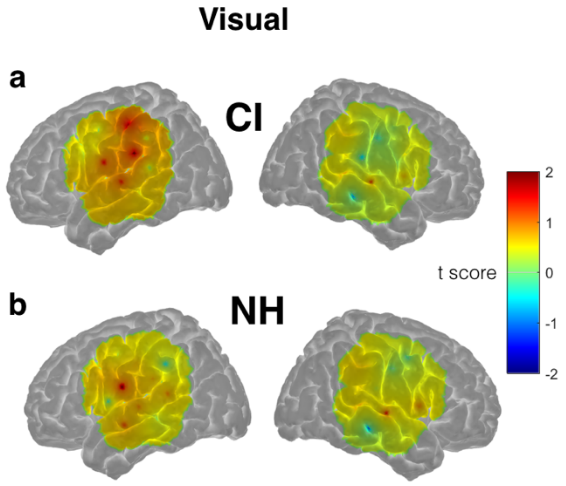


FIGURE 4.10. Cortical activity during visual only lipreading. No significant differences were found in either CI (a) or NH (b) groups.

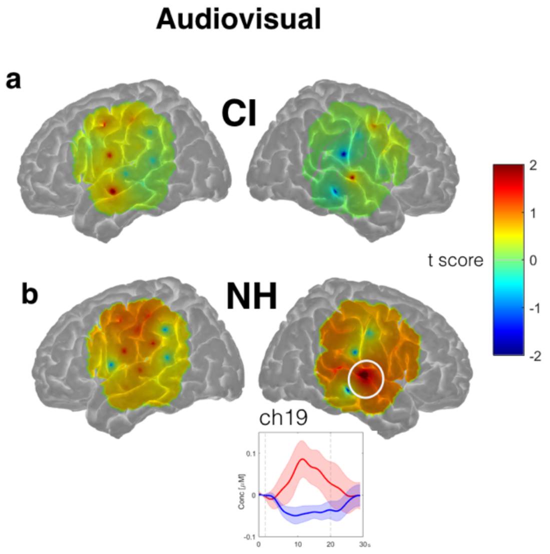


FIGURE 4.11. Cortical activity during audiovisual listening. No significant channels were identified in the CI group (a), while a right-lateralized auditory-association area had a high peak concentration of HbO at 0.91 μM in the NH group (b).

Task	Metric	p(McGurk) correlation Kendall's tau (p value)	
		CI	NH
Word recognition	Lipreading average	0.0006 (0.97)	0.16 (0.30)
	High SNR ii	-0.18 (0.29)	0.02 (0.89)
	Med SNR ii	0.087 (0.62)	-0.15 (0.36)
	Low SNR ii	-0.24 (0.15)	-0.09 (0.57)
fNIRS	Auditory ch 11 β	0.029 (0.88)	-0.07 (0.67)
	Auditory ch 22 β	0.13 (0.52)	**0.38* (0.015)**
	Auditory ch 8 β	n/a	**0.35* (0.023)**
	Auditory ch 19 β	n/a	0.30 (0.054)
	Audiovisual ch 19 β	n/a	0.13 (0.42)

TABLE 4.5. Correlations with the McGurk effect.

Neither the CI group, nor the NH group had any significant differences within the region of interest during the lipreading condition, and activation patterns overall appeared quite similar between groups (Fig. 4.10). Therefore, no crossmodal plasticity of visually-driven auditory cortex was detected in the CI group. Similarly, in audiovisual blocks no channels in the CI group reached significance (Fig. 4.11a); however, the NH group did display strong activity in the right temporal lobe near auditory cortex (Fig. 4.11b), an area corresponding to channel 19 (Table 4.4). Finally, in the written condition there was seemingly higher activity across all channels, and resulting in significant activity in the left hemisphere of the NH group at channel 8 (Fig. 4.12b).

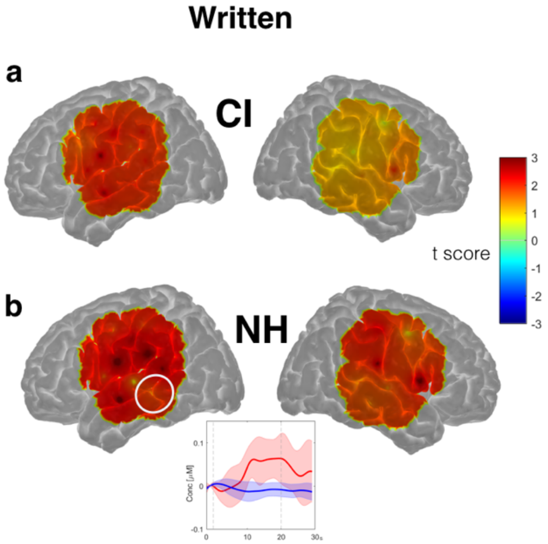


FIGURE 4.12. Cortical activity during the written condition. CI users (a) had no significant channels, yet the NH group (b) was significant at channel 8.

Correlations between tasks. Next, we asked whether the behavioral measures of AV processing were correlated to one another and/or to the aforementioned functional measures. For both groups, we found that the probability of McGurk fusion neither correlated to lipreading proficiency nor to any measures of interactive index (ii) at any SNR level (Table 4.5). However, two significant recording channels during auditory-only fNIRS blocks *were* significantly correlated with p(McGurk) in the normal hearing group. This relationship was not present in the CI group (Table 4.5).

Task	Metric	Clinical measures in CI users CNC quiet	CNC +10 dB	CNC +5 dB	Hearing Loss [y]	Deafness [y]	CI use [y]
Word recognition	High SNR ii	-0.18 (0.23)	-0.11 (0.56)	-0.07 (0.72)	**0.35* (0.02)**	0.22 (0.16)	0.24 (0.11)
	Med SNR ii	-0.11 (0.53)	0.009 (0.96)	-0.05 (0.79)	0.02 (0.89)	0.21 (0.24)	0.32 (0.06)
	Low SNR ii	-0.19 (0.23)	0.008 (0.97)	0.09 (0.65)	-0.19 (0.22)	0.26 (0.11)	0 (1)
	Lipreading average	0.05 (0.75)	**-0.43* (0.02)**	**-0.40* (0.012)**	0.17 (0.28)	-0.11 (0.50)	0.16 (0.31)
fNIRS	Auditory ch 11 β	0.07 (0.70)	0.37 (0.06)	**0.54** (0.008)**	0.09 (0.60)	0.07 (0.70)	0.24 (0.15)
	Auditory ch 22 β	0.27 (0.11)	0.25 (0.20)	0.25 (0.23)	-0.07 (0.67)	0.17 (0.32)	0.02 (0.89)
McGurk	p(McGurk)	-0.07 (0.70)	0.019 (0.92)	0.117 (0.55)	0.19 (0.26)	0.15 (0.37)	-0.07 (0.65)

TABLE 4.6. Correlations with clinical measures within the CI group.

Clinical correlations within the CI group. Finally, we asked whether clinical measures within the CI group correlated with AV integration at the whole world level, the McGurk task, lipreading ability, or fNIRS metrics (Table 4.6). We found that the duration of hearing loss was positively correlated with AV gain in quiet (i.e., "high" SNR ii). Additionally, clinical word testing in noise had a negative relationship with lipreading skill (Table 4.6). In other words, better lipreaders scored more poorly on auditory-only listening in noise tests. Finally, the most difficult CNC testing at a +5 dB SNR, for which 14 CI users also had fNIRS data, we found a positive correlation with cortical activity at channel 11 near speech processing in the vicinity of the supramarginal gyrus and Brodmanns area 40 (Table 4.4).

4.4 Discussion

Despite being better lipreaders, we did not see any corresponding increases in AV word recognition or ii for the CI group (Fig. 4.6). We did, however replicate prior findings of a visual bias in the McGurk task and found that CI users are also less likely to fuse McGurk stimuli into a novel percept (Fig. 4.8d). This difference in p(McGurk), however, was only evident after correcting for misidentifications of unisensory stimuli for the fused percept, suggesting that auditory ambiguity for CI users should be taken into account in future studies calculating the magnitude of illusion perception.

As expected, CI users who did not perceive the illusion, overwhelmingly reported the visual component of the incongruent trials (i.e., the viseme ‘ga’). In contrast, NH controls who did not perceive the illusion were much more likely to report the auditory stimulus. This discrepancy between groups may suggest CI users “weight” the visual component of natural speech higher than typical hearing populations. Despite these clear differences between groups in the McGurk task, they seem to have little consequence for CI users given that likelihood of fusing syllables doesn’t correlate with any clinical variables, lipreading skill, or measures of AV integration (Table 4.5).

Because the McGurk task is an artificial circumstance of incongruent information that lacks any semantic content, we asked whether simulating more realistic listening environments would reveal difference in integration. Background noise reduces the saliency of auditory information and successful discourse requires integration with complementary visual speech information from orofacial articulations. We sought to match auditory performance between groups and to have similar, real-world demands in both groups—a perspective that let us *not* to vocode speech for the NH group, but instead use louder babble for controls. We found highly-similar results between groups in this task for AV listening and multiple measures of AV gain (ii and visual enhancement). These results suggest that CI users may integrate semantic speech similar to their NH counterparts. The possible utility of increased visual weighting is evident in the CI user visual modality having the highest performance in the word recognition task (Fig. 4.7). Thus, higher lipreading performance seems to effectively compensate for auditory deficiencies such that AV gain is equivalent between groups. In summary, CI users utilize visual inputs to a greater extent in order to integrate auditory and visual speech just as well as controls.

It should also be noted that although AV integration is the same between groups, NH controls are tested in an average of 64 dB of noise in the easiest condition, which approximates listening ability in quiet for CI users. In contrast, listening in quiet for the NH group would result in performance at ceiling and no experience of AV benefit. For CI users, no listening condition is too easy for visual articulations to not be beneficial. This finding underscores the importance of better understanding the mechanisms of AV processing, because CI users are effectively employing this technique all the time.

We did not identify crossmodal plasticity of visually-driven auditory cortex in the CI group as others have indicated (Giraud et al., 2001; Stropahl and Debener, 2017). In fact, we did not see any significant activity in either group during lipreading (Fig. 4.10). It is possible that our optode holder was situated too anterior to detect visual activity during lipreading (as in Chapter 3). Unfortunately, CI receivers are so close to these anatomical areas, it's unclear whether they could be imaged even with more dense recordings. Potentially including a variety of smaller and overlapping channels would improve spatial resolution to further explore how visual input from lipreading is processed in CI users' brains.

We did detect some auditory-evoked activity in the CI group, but surprisingly, we did not find any significant channels in the AV listening condition (Fig. 4.11). It's possible that these areas were too close to implants for reliable recordings. Because we did see some AV activity in the right PAC for the NH group, the task design does seem capable of capturing this multimodal activity. It is possible that cortical activity in this condition was more diffuse, and/or the probe locations were not consistently placed. As a result, highly-localized activity may not be evident at the group level. Future work using a 3D digitizer would help confirm that probe registration is consistent between individuals.

Two channels bilaterally had a positive correlation with McGurk perception, but only in the control group (Table 4.5). While these are near cortical areas of multisensory integration, and it is feasible that their activation is causally related to the illusion percept, further work employing a causal research design is necessary. For instance, reversibly deactivating the left STS via transcranial magnetic stimulation (TMS) has been shown to reduce illusion perception by as much as 50% (Beauchamp et al., 2010). These experiments would be particularly interesting to contrast the effects of deactivation in CI groups who may be affected more or less than NH controls.

Additionally, within the CI group, there are negative correlations between CNC testing in noise and lipreading performance (Table 4.6). This suggests that lipreading proficiency may increase adaptively as listening-in-noise is more challenging for some individuals. Interestingly, a positive correlation between auditory-evoked left hemisphere activity and CNC testing in noise suggests that more proficient CI users have greater left hemispheric auditory processing—a finding similar to what others have reported (Lazard et al., 2010; Lee et al., 2001).

In conclusion, CI users clearly experience visually-biased perception of speech that seems to fully compensate for auditory listening difficulties and result in magnitudes of AV integration that's indistinguishable from controls. The proficiency in lipreading that underlies this skill is more pronounced in CI users who have poor clinical speech-in-noise outcomes. In general, CI users are better lipreaders than controls, and further work is required to better understand the neural processes responsible for this advantageous behavior. At present, it seems that left lateralized auditory-evoked activity relates to higher speech outcomes, though we did not identify indicators of crossmodal plasticity or visually-evoked differences.

4.5 Acknowledgments

Laura Levin and Linsey Sunderhaus recruited CI users. Andie DeFreese recruited NH participants and assisted in data collection. Iliza Butera designed the experiments, collected data, analyzed the data, and wrote the manuscript. Jasmine Greer and Sharae Cockrill provided technical support for fNIRS acquisition. Baxter Rogers provided feedback on the fNIRS experimental design and analysis. Mark Wallace and René Gifford provided guidance and feedback throughout the experiments and manuscript preparation.

CHAPTER 5 Audiovisual temporal processing in postlingually deafened adults with cochlear implants

5.1 Introduction

Cochlear implantation is an effective surgical intervention for individuals with severe-to-profound sensorineural hearing loss to either regain auditory speech perception or, for the congenitally deaf, to establish it for the first time. This highly successful neuroprosthetic device parcels acoustic signals into frequency bins that correspond to tonotopic stimulation of intracochlear electrodes. Despite considerable technological and surgical advancements, spread of electrical excitation in the cochlea remains a significant barrier for cochlear implant (CI) users to achieve high-fidelity spectral encoding. As a result, the degraded auditory signal that an implant provides can be quite ambiguous to some CI users (Gifford et al., 2008, 2014; Holden et al., 2013).

Fortunately, speech is typically an audiovisual (AV) experience wherein coincident orofacial articulations can considerably boost perceptual accuracy over that observed with auditory-alone stimulation (Sumby and Pollack, 1954). Indeed, a great deal of modeling work suggests that ambiguous information stemming from unreliable sensory estimates is optimally integrated in the brain by weighting the relative reliability of the different sources of sensory evidence (Burr and Alais, 2006; van Dam et al., 2014; Ernst and Bülthoff, 2004). This process results in a more robust multisensory percept with specific advantages including increased stimulus saliency (Stein et al., 1996), decreased detection thresholds (Grant and Seitz, 2000b; Lovelace et al., 2003b), reduced reaction times (Gilley et al., 2010), and enhanced efficiency in neural processing (Van Wassenhove et al., 2005).

Many of these multisensory-mediated benefits have been seen in children who have received early cochlear implantation (i.e., before age 4). These include: faster reaction times (Gilley et al., 2010), greater multisensory gain (Bergeson et al., 2005; Lachs et al., 2001), and higher speech recognition at multiple levels of phonetic processing (Tyler et al., 1997). Furthermore, it has been suggested that many CI users may achieve audiovisual speech recognition abilities that are comparable to normal-hearing individuals after matching unisensory performance (e.g., through masking or generating CI simulations of speech for typical listeners) (Rouger et al., 2007). In support of this claim, several studies in both CI users and other hearing-impaired populations indicate proficient multisensory integration as measured via AV speech recognition tests of consonants (Tye-Murray et al., 2007), phonemes (Desai et al., 2008b), words (Kaiser et al., 2003; Rouger et al., 2007;

Schreitmüller et al., 2017), and sentences (Lachs et al., 2001; Moody-Antonio et al., 2005), as well as in work that employs computational models (Grant et al., 1998; Massaro and Cohen, 1999).

The perceived timing of auditory and visual information is a core determinant in the efficacy by which cues are integrated and perceptually 'bound' (see below) (Burr and Alais, 2006; Freeman et al., 2013; Vroomen and Keetels, 2010). Although the aforementioned studies illustrate that AV integration can be quite good in CI users, temporal processing is a critical factor in this process and, somewhat surprisingly, very little is known about AV temporal processing in the CI population. Thus, we questioned whether the auditory information conveyed by a cochlear implant may alter the perceptual weighting of visual and auditory cues in such a way that generalizes to differences in AV temporal assessments compared to normal-hearing individuals.

In typical development, it is well established that AV stimuli are more likely to be perceptually bound when the individual component stimuli are in temporal (as well as spatial) proximity (Vroomen and Keetels, 2010). However, AV stimuli need not be precisely synchronous for this binding to occur, but rather appear to be integrated over a range of temporal intervals spanning several hundred milliseconds, a construct known as the temporal binding window (TBW) (Wallace and Stevenson, 2014). Only one published study has investigated the temporal binding of AV stimuli in CI users, and the authors indicated no difference in TBWs between age-matched, normal-hearing (NH) individuals and CI users during the presentation of monosyllabic words (Hay-McCutcheon et al., 2009). Other work has reported similar findings for the moderately hearing-impaired while judging whether AV sentences were either synchronous or asynchronous (Baskent and Bazo, 2011). In contrast, few studies have investigated AV temporal function across a broader range of stimulus types ranging from the simplistic (i.e., flashes and beeps) to the more speech-related (Gilley et al., 2010; Gori et al., 2017; Stevenson et al., 2017a). Furthermore, given the use of word and sentence stimuli in prior work, we sought to examine here whether differences in AV temporal performance would be evident in CI users while making less complex, sublexical temporal judgments.

The present study investigates multisensory (i.e., combined audiovisual) temporal processing in CI users and a group of NH controls using speech syllables and simple "flashbeep" stimuli. Because early access to sound is an important factor for the development of multisensory integration, we recruited postlingually deafened adult CI users to test whether auditory, visual, and audiovisual temporal functions are altered in those who experience typical auditory development in early life (in contrast to prelingual deafness). In normal-hearing individuals, the TBW has been shown to both narrow as well as become more asymmetrical during development (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012). Thus, our primary hypothesis was that CI users, given their altered AV experience, would exhibit broader temporal binding windows centered closer to objective simultaneity (i.e., 0 ms) than controls. Practically, this would mean CI users are less able to accurately identify AV asynchronies. We expected this result to be specific for speech stimuli and not for simple flashbeep stimuli on account of the greater ecological validity of speech signals. We drew this prediction in part from prior work investigating the maturation of temporal binding windows in normal development (Hillock-Dunn and Wallace, 2012), and reasoned that reduced auditory experience during deafness might result in less mature (i.e., broader and more symmetric) temporal binding windows to be evident well into adulthood for CI users.

We tested our hypothesis using two distinct tasks: simultaneity judgment (SJ) and temporal order judgment (TOJ)—the latter of which we also used to quantify unisensory temporal thresholds (Fig. 5.1). During SJ tasks, which are commonly used to measure TBWs, individuals are presented with auditory and visual stimuli that vary in relative synchrony and asked to report whether they perceived the two stimuli to have occurred at the "same time" or at "different times." SJ tasks were administered using both a simple (i.e., "flashbeep") stimulus and a more complex speech syllable stimulus presented at 12-19 different stimulus onset asynchronies (SOAs) (see Fig. 5.1). In a common measure closely tied to TBW, we also quantified the point of subjective simultaneity (PSS)—the SOA at which maximal reports of perceived synchrony occurred (Fig. 5.2a). In adults, PSS is typically visual leading (by convention represented as a positive SOA), which likely reflects an adaptation to the relative physical transmission speeds of light versus sound (Vroomen and Keetels, 2010).

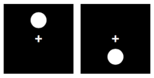
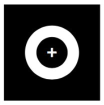
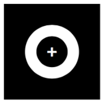
Task	SOAs [ms]	Trials per SOA (total)	Stimuli		Prompt	Time [min]
			Visual	Auditory		
Visual TOJ	±10, ±20, ±30, ±40, ±60, ±80, ±100, ±150	10 (160)		–	If the top flash came first, press 1. If the bottom flash came first, press 2.	5
Auditory TOJ	±10, ±20, ±35, ±50, ±75, ±100, ±150, ±200, ±250	10 (180)		500 Hz tone 2 kHz tone	If the high pitch came first, press 1. If the low pitch came first, press 2.	7
SJ speech	0, ±50, ±100, ±150, ±200, ±250, ±300, ±400	20 (300)		 "ba"	If the audio and the video were properly lined up, press 1. If the audio and the video were NOT properly lined up, press 2.	20
SJ flashbeep	0, ±10, ±20, ±50, ±80, ±100, ±150, ±200, ±250, ±300	20 (380)		1 kHz tone	If the flash and the beep were at the same time, press 1. If the flash and the beep were at different times, press 2.	13
AV TOJ speech	±50, ±100, ±150, ±200, ±250, ±300, ±400	20 (280)		 "ba"	If the visual came too early, press 1. If the audio came too early, press 2.	20
AV TOJ flashbeep	±20, ±50, ±80, ±100, ±150, ±200, ±250, ±300	20 (320)		1 kHz tone	If the visual flash came first, press 1. If the auditory beep came first, press 2.	13

FIGURE 5.1. Unisensory and multisensory psychophysical tasks. Both temporal order judgment (TOJ) and simultaneity judgment (SJ) tasks utilized either circles and tones or speech. SOA= stimulus onset asynchrony.

Additionally, an AV TOJ task utilized the same stimuli as the SJ task but with instructions to report the stimulus order instead of the apparent synchrony. Finally, unisensory temporal processing was also assessed with auditory TOJ (aTOJ) and visual TOJ (vTOJ) tasks wherein two brief unisensory stimuli (e.g., two circles or two tones) are presented in rapid succession at varying SOAs, and individuals report which stimulus occurred first. In prior studies from our group, this testing battery has been used to evaluate temporal thresholds of unisensory and multisensory processing of simple stimuli and speech syllables in typical populations across the lifespan (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012; Stevenson et al., 2017b) and in individuals with neurodevelopmental disorders (Stevenson et al., 2014c, 2017c; Woynaroski et al., 2013). It is employed here using similar methods to evaluate adult postlingually deafened CI users in comparison to NH controls.

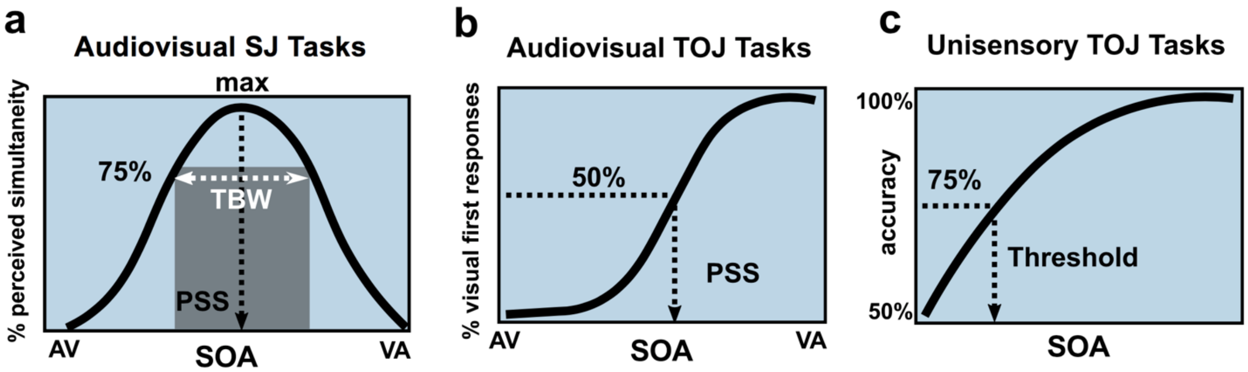


FIGURE 5.2. Summary metrics for each temporal task. For SJ tasks (**A**), the percent of perceived simultaneity is plotted per SOA from auditory-leading to visually-leading offsets in order to derive TBW at 75% and the PSS from the peak of the function. Two opposing logit functions are used for TBW curve fits so that symmetry is not assumed, while Gaussian functions are used for PSS derivation. For AV TOJ tasks (B), the percent of visual-first responses is plotted, and PSS at 50% is calculated from the resulting sigmoid functions. For unisensory TOJ tasks, percent accuracy is plotted in order to collapse across positive and negative SOAs that are arbitrarily defined as the top circle/high pitch occurring first (+SOA) or the bottom circle/low pitch occurring first (-SOA). Threshold is derived from a logit function at 75% accuracy. SJ = simultaneity judgement; TBW = temporal binding window; SOA = stimulus onset asynchrony; PSS = point of subjective simultaneity; TOJ = temporal order judgment; AV = audiovisual

5.2 Methods

Participants. This study included 56 postlingually deafened CI users and 55 NH controls between the ages of 19 and 77 years old (Table 5.1). Four participants (3 CI users, 1 NH control) were excluded from final analyses due to excessive missing data (i.e., for more than 50% of the tasks). Five additional participants (all CI users) were excluded due to: non-functional implants (n= 2), impaired vision (n= 1), and other confounding neurological diagnoses (n= 2). On average the NH controls (N = 54) were 8.4 years younger than CI users (N = 48; $t_{(1,100)}$= 2.8, *p*=0.007). As a result, age was included as a covariate for between-groups comparisons (see Results).

Group	N	Sex (% female)	Mean age ± SD (y)	PTA (dB SPL)		Number of CIs	Acoustic hearing	Implant manufacturer	nonverbal IQ
				L	R				
CI	48	54%	53.4± 13.6*	26 ± 7	25 ± 6	1 – 60% 2 – 40%	48%	60% Cochlear 29% MED-EL 10% AB	102 ± 15*
NH	54	76%	45.0± 16.5	9 ± 7	10 ± 8	—	100%	—	109 ± 13

TABLE 5.1. Clinical characterization of cochlear implanted and normal hearing groups. Pure Tone Averages (PTA) measure aided detection thresholds for the CI group; values are means ± standard deviation, * p<0.05 CI = cochlear implant; NH= normal hearing; y= years; PTA = pure tone averages; dB SPL = decibels of sound pressure level; AB = Advanced Bionics

In the control group, standard audiometric testing ensured normal pure tone averages (left ear 9±7 dB, right ear 10±8 dB) and speech perception (AzBio sentences, range 98-100% correct). We obtained aided audiograms at the study visit for the majority of CI users (n= 27) indicating that pure tone averages were appropriate for stimulus audibility (left 25±6 dB, right 25±6 dB). The remaining CI users (n=21) were screened for detection at 30 dB at the time of testing. Speech perception in the CI group was measured via standard monosyllabic, consonant-nucleus-consonant (CNC) word lists, which indicated a wide range of proficiency consistent with other reports in the literature (Gifford et al., 2008).

CI users were required to have at least 3 months of experience with their implants prior to testing. The average experience was 3.5 years (range of 4 months to 11 years). The mini-mental state exam (MMSE) was also administered to screen for cognitive impairment defined by scores below 24 (of 30 possible points), and no exclusions were made based on this criterion (CI 29 ± 2; NH 29 ± 3). To minimize possible confounds with language, we focused on the nonverbal subscore of the Kaufman Brief Intelligence Test (KBIT) that indicated both groups had mean scores of nonverbal cognition within the age-normative range. There were, however, significant group differences where the controls scored slightly higher on average (CI = 102 ±15; NH= 109 ±13; $t_{(87)=}$ -2.1, p=0.038).

All CI users were postlingually deafened and used the most current generation sound processors at the time of experimentation for all auditory testing. Additionally, all participants were screened for visual acuity by verbal confirmation and/or a Snellen eye chart, wearing corrective lenses as needed.

Stimuli. Visual stimuli were generated in Matlab 2008a using Psychtoolbox extensions (Kleiner et al., 2007). They were presented on a CRT monitor (100 Hz refresh rate) positioned approximately 50 cm from participants. Visual stimuli were white rings and circles (10 ms in duration) on a black background. Articulations of the syllables “ba” and “ga” were produced by an adult female speaking at a normal rate and volume with a neutral facial expression. Auditory stimuli were delivered at a comfortably loud level (calibrated to 65 dB SPL in the sound field) presented through stereo speakers. Auditory stimuli included tones (10 ms in duration) ranging from 500 Hz to 2 kHz as well as utterances of the syllables “ba” and “ga.” For flashbeep tasks, an oscilloscope (Hameg Instruments, HM407-2) was used to align auditory and visual stimuli to objective synchrony (0 ms) or the stimulus onset asynchronies (SOAs) shown in Figure 5.1. For speech tasks, natural speech was considered objective synchrony.

Procedures. All protocols and procedures were approved by Vanderbilt University Medical Center's Institutional Review Board, and all methods were performed in accordance with these guidelines and regulations, which included all volunteers providing informed consent prior to participation. Experiments took place in a dimly lit, sound-attenuated room with an experimenter seated nearby. Both task order and trial order were pseudo randomized. On all non-speech tasks, subjects were instructed to maintain fixation on the centrally-located fixation cross. All responses were collected using a standard keyboard, and all testing was completed over 1 or 2 study visits.

Analysis. For the unisensory TOJ tasks, we collapsed across positive and negative SOAs to plot accuracy at each temporal offset regardless of the stimulus position (for vTOJ) or frequency (for aTOJ). These data points were then fit with a standard logit function using the Matlab function glmfit to derive a threshold at 75% of the psychometric curve. For all tasks, any subject's threshold that exceeded the largest SOA was excluded. Individuals who had accuracy higher than 75% at all tested SOA were assigned a conservative threshold of the smallest SOA tested, which was 10 ms for both aTOJ and vTOJ tasks.

In all multisensory tasks, negative SOAs correspond to auditory-leading stimuli and positive SOAs correspond to visually-leading stimuli. Any bias in PSS is indicated by the sign, which typically reflects perceptual biases related to the slightly visually-leading onsets of natural speech (i.e., delayed voicing relative to orofacial articulations) and adaptations to differing physical transmission times of light and sound (Vroomen and Keetels, 2010). Because objective synchrony at 0 ms is the only point at which participants are truly "correct" in their simultaneity judgment, these tasks are unsuitable for further SDT analysis, so we focused this analysis on TOJ tasks where responses could be coded as hits (H) and false alarms (F).

We calculated d′ using the conventional formula of subtracting the z scores of the false alarm rate from the hit rate ($d' = z(H) - z(F)$), with hits defined as correct "visual-first" responses and false alarms defined as incorrect "visual-first" responses. Put another way, for the AV TOJ tasks, as an example, we coded responses as if in response to the question "did the visual stimulus appear first?" such that "hits" were "yes" responses to positive SOAs (i.e., VA) and "false alarms" were "yes" responses to negative SOAs (i.e., AV). Next, we calculated a measure of response bias using the formula c = -0.5 x [z(H) + z(FA)]. It should be noted that in this circumstance, response bias is not a fixed characteristic of the observer, but instead shifts systematically with SOA. This results from the fact that we are not calculating criterion from a 0 ms SOA, which would indicate an individual's fixed criterion. Instead, as the magnitude of the SOA increases, the proportion of correct "visual-first" responses, for instance, necessarily increases and the proportion of incorrect "auditory first" responses necessarily decreases. Confidence intervals at each SOA were calculated as discussed in (Macmillan and Creelman, 2004).

For SJ tasks, mean reports of synchrony were plotted at each SOA, and the data were fit with two intersecting logit functions. All curves were normalized to 100% perceived simultaneity for TBW calculations, which was defined as the distance between the left and right curves at 75% reported synchrony. Simultaneity data were also fit with single Gaussian functions in R software (R code team, 2012) to derive the PSS at the peak of each curve.

SPSS Statistics for Macintosh Version 24.0 (IBM) and Prism 7.0b for Mac OSX (Graphpad Software) were used for statistical comparisons and graphing, respectively. All figures illustrate 95% confidence intervals of the mean. Otherwise, variance is indicated as standard deviation throughout the text.

Statistical approaches. We utilized a non-parametric approach to evaluate between-group differences in audiovisual integration (i.e., bootstrapping). Missing data occurred for several reasons including insufficient testing time and more commonly, the inability to derive thresholds from curves for a variety of reasons such as participant fatigue, poor attention, misunderstood instructions, and insufficient SOA magnitudes. Missing data was handled by pairwise deletion.

Univariate regressions were carried out in initial tests of between-group differences in temporal processing across eight summary metrics (SJ speech PSS, SJ flashbeep PSS, SJ speech TBW, SJ speech PSS, aTOJ threshold, and vTOJ threshold, avTOJ flashbeep PSS, and avTOJ speech PSS). Given aforementioned between-group differences in age and nonverbal IQ, these background variables were explored as covariates and retained in all models where they accounted for significant variance. Multivariate follow-up analyses were used to further characterize the nature of statistically significant between-group differences. When significant, further univariate test of SOA-level differences were performed.

5.3 Results

Using nonparametric (i.e., bootstrapped) univariate regressions, we performed planned analyses of between-group differences across eight summary metrics (Fig. 5.2) derived from six psychophysics tasks. Group differences in age and nonverbal IQ prompted us to explore these indices as covariates (see Methods for participant characteristics); these factors were retained in all models wherein they accounted for significant variance. To further explore significant findings, post hoc follow-up tests evaluated between-group differences at all individual SOAs using multivariate analysis of variance (MANOVA) or a multivariate analysis of covariance (MANCOVA), where appropriate.

Psychophysical tasks	F statistic	Significance	Effect size (partial η^2)
SJ flashbeep PSS	$F_{(1,91)} = 1.9$	p = 0.17	0.021
SJ flashbeep TBW	$F_{(1,85)} = 0.47$	p = 0.49	0.006
SJ speech PSS	$F_{(1,97)} = 8.8$	**p = 0.004**	0.084
SJ speech TBW	$F_{(1,88)} = 0.96$	p = 0.33	0.011
Visual TOJ threshold*	$F_{(1,82)} = 8.6$	**p = 0.004**	0.095
Auditory TOJ threshold	$F_{(1,77)} = 1.2$	p = 0.28	0.015
avTOJ flashbeep PSS	$F_{(1,74)} = 0.33$	p = 0.57	0.004
avTOJ speech PSS	$F_{(1,57)} = 3.7$	p = 0.06	0.060

TABLE 5.2. Between-subjects univariate regressions for temporal tasks. Covariates included where indicated by an asterisk. Significance was set at $\alpha = 0.05$, and bolded values met this threshold. TOJ = temporal order judgment; SJ = simultaneity judgment, TBW = temporal binding window; PSS = point of subjective simultaneity.

Overview of findings for summary metrics. A statistically significant difference in performance was found between groups for SJ speech PSS, with mean values of 15.5 ms in the CI group and 54.7 ms in the NH controls (p = 0.004, Table 5.2). Additionally, a significant between-group difference was observed in vTOJ thresholds when controlling for age and nonverbal IQ (which were significant predictors in the vTOJ regression model). Interestingly, CI users had improved thresholds relative to NH controls (corrected means = 38.1 ms and 49.5 ms, respectively; p = 0.004).

Thus, the CI group has: 1) PSS for speech that is shifted *away* from visually-leading SOAs (i.e., less positive) and 2) improved visual temporal thresholds. Between-group differences were not observed for any other measures; however, PSS for the audiovisual speech TOJ task was noteworthy in its marginally significant group differences (p = 0.057, Table 5.2) between CI users and controls (-78.2 ms and -7.83 ms, respective means), which is consistent with the PSS shift observed in response to the SJ speech task.

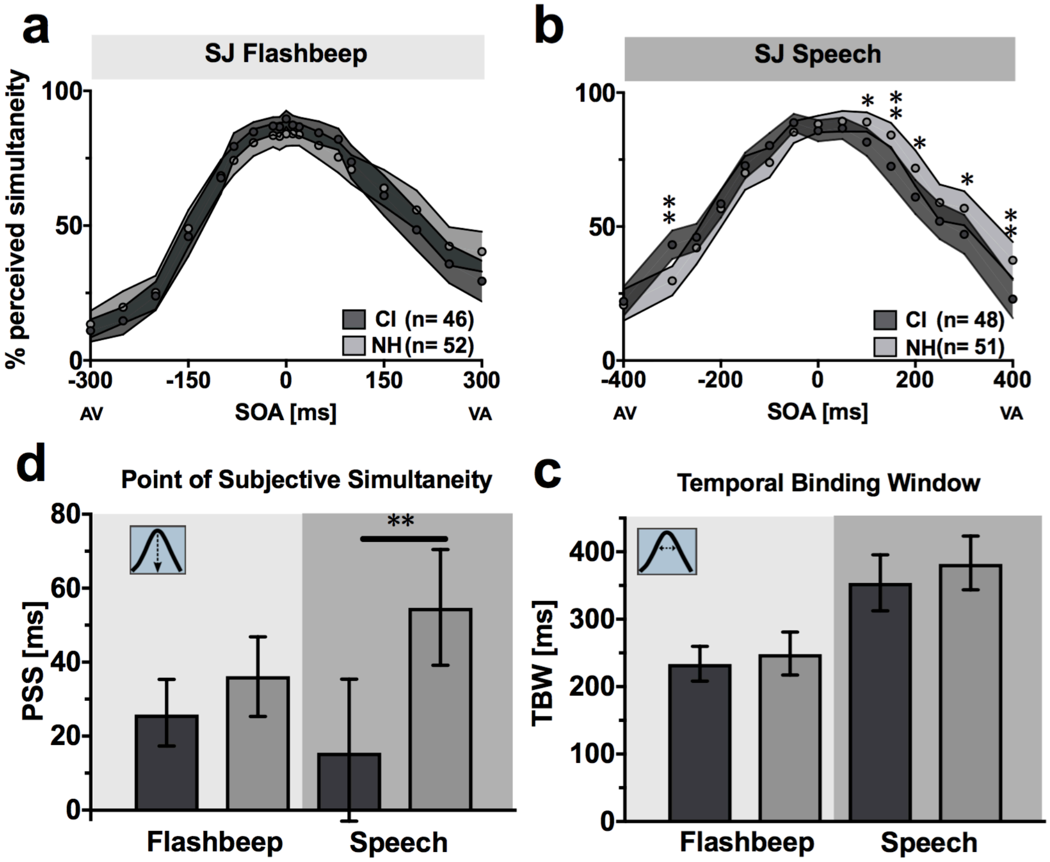


FIGURE 5.3. Simultaneity judgment tasks. Mean reports of simultaneity at each SOA are plotted for flashbeep (A) and speech tasks (B). CI users are shown in dark gray and NH in light gray. Circles indicate means and shaded areas correspond to 95% confidence interval of the mean. Mean PSS (C) and TBW (D) calculations are shown with error bars also indicating 95% confidence intervals. CI = cochlear implant; NH = normal hearing; SOA = stimulus onset asynchrony; TBW = temporal binding window; PSS = point of subjective simultaneity; *p<0.05, **p<0.01

Perceived synchrony of speech stimuli is less visually leading in CI users. Audiovisual temporal function was examined using the SJ task wherein we varied the timing onset between the auditory and visual components of either a syllable/viseme or a less complex circle/beep pairing and had participants indicate if the pair appeared synchronous. As illustrated in Figures 5.3a and 5.3b, performance is plotted as mean reports of synchrony for each SOA. Not surprisingly, for the flashbeep task (Fig. 5.3a), confidence intervals are highly overlapping between CI users and NH controls. This plot further supports the aforementioned lack of group difference in AV temporal acuity for low-level stimuli (Table 5.2), which requires no further follow-up testing. In contrast, a post hoc MANOVA for the speech task (dependent variables: 15 SOAs) indicated a statistically significant difference between groups ($F_{(15,83)}$= 3.05, p= 0.001; Wilk's Λ = 0.645, η_p^2= 0.355). Follow-up univariate tests indicated that the CI group differed from the NH controls at six (out of 15) SOAs. These include the following negative (i.e., auditory leading) asynchronies and positive (i.e., visually-leading) asynchronies: -300 ms ($F_{(1,97)}$= 11.3, p =0.001, η_p^2=0.11), 100 ms ($F_{(1,97)}$= 8.009, p =0.006, η_p^2=0.076), 150 ms ($F_{(1,97)}$= 9.66, p =0.002, η_p^2=0.091), 200 ms ($F_{(1,97)}$= 5.62, p =0.02, η_p^2=0.055), 300 ms ($F_{(1,97)}$ = 4.49, p =0.037, η_p^2=0.044), and 400 ms ($F_{(1,97)}$ = 9.84, p =0.002, η_p^2=0.092). These differences appear to be a result of group-level differences in PSS, which was derived from individual curve fits then averaged across groups (Fig. 5.2a). Thus, the overall shift to the left (toward 0 ms) in CI users relative to NH subjects is evident both in the individual PSS calculations (Fig. 5.3c) and the averaged responses at each SOA (Fig. 5.3b).

Contrary to our hypothesis, there were no significant differences in TBW for either the flashbeep or speech stimuli (Table 5.2; Fig. 5.3d). Collectively, these results indicate a shift in AV temporal performance in CI users that is specific for speech stimuli, and that manifests not as a change in overall temporal acuity (i.e., a TBW shift), but rather as a shift in the peak of this Gaussian function toward objective synchrony (i.e., a PSS shift toward an SOA of 0 ms).

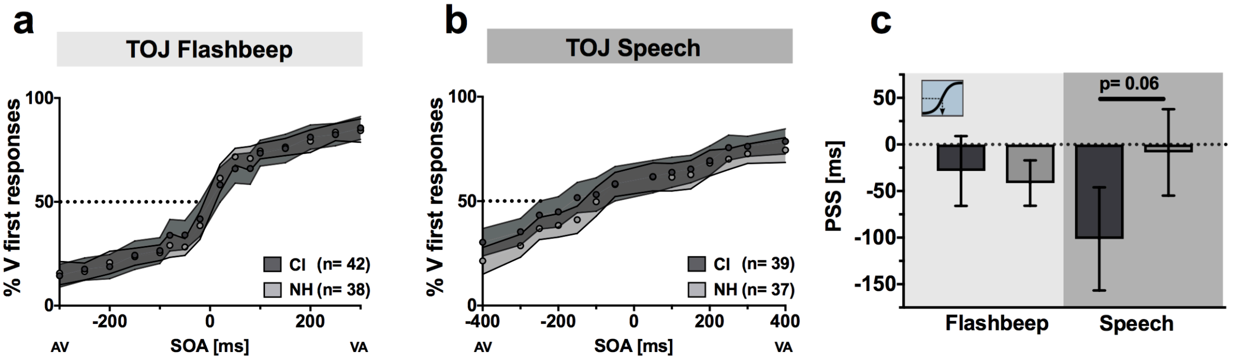


FIGURE 5.4. Audiovisual temporal order judgment tasks. Mean percent of "visual-first" responses per SOA are plotted for audiovisual flashbeep (A) and speech (B) TOJ tasks. CI users are shown in dark gray and NH in light gray. Circles indicate means, and shaded areas correspond to 95% confidence intervals of the mean. PSS (point of subjective simultaneity; C) for both tasks were derived for each subject and averaged across groups by the intersection of the psychometric function with 50%. Error bars indicate 95% confidence intervals. CI = cochlear implant; NH= normal hearing; SOA= stimulus onset asynchrony.

TOJ of multisensory speech further supports a shift in PSS between groups. In order to further explore the shift in PSS for CI users, we tested AV temporal function using AV TOJ tasks. With the same stimuli from the SJ tasks, individuals were instructed to indicate whether the visual or auditory component occurred first. The proportion of "visual first" responses is plotted for each SOA in the flashbeep (Fig. 5.4a) and speech task (Fig. 5.4b). The PSS for TOJ is calculated as the SOA at which the two alternatives (auditory first or visual first) are equally likely. It should be noted that maximum perception of synchrony (SJ PSS) is similar but not equivalent to the maximum uncertainty of the presentation order (TOJ PSS) (Eijk et al., 2008; Love et al., 2013; Vatakis and Spence, 2006), because it is possible to perceive two objects as asynchronous but to still be unsure of which occurred first. Accordingly, the PSS values for the TOJ tasks (Fig. 5.4c) differ in magnitude compared to the SJ tasks (Fig. 5.3c), yet appear to support the result that CI users differ for PSS only with the speech tasks. Although temporally shifted judgments of speech TOJ were not statistically significant (Table 5.2), the overall pattern is consistent with the SJ task and more importantly, also provide an opportunity to further investigate underlying decisional biases in AV temporal judgments using signal detection theory (SDT) derived analyses.

Signal detection analysis of the audiovisual speech TOJ task reveals a visual response bias in CI users. After pooling responses from all subjects in each group for all four TOJ tasks, we applied signal detection methods to calculate measures of sensitivity (d′) and response bias (c) (see Methods). For these measures, larger d′ values correspond to higher sensitivity or discriminability, and bias values of zero represent an "unbiased" response (i.e., one in which the criterion is unchanged). For the AV TOJ tasks, we calculated the probability of correct "visual first" responses to negative SOAs (i.e., VA trials) and incorrect "visual first" response to positive SOAs (i.e., AV trials). Non-overlapping confidence intervals are considered significant group differences at the corresponding SOAs for both sensitivity and bias.

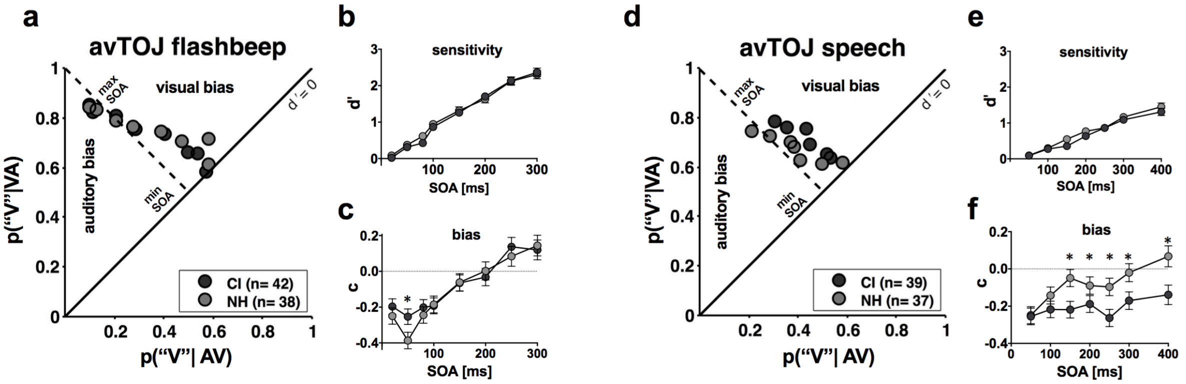


FIGURE 5.5. Signal detection analysis of audiovisual temporal order judgment tasks. The probability of hits versus probability of false alarms are plotted for the audiovisual TOJ flashbeep (A) and audiovisual TOJ speech tasks (D). From these ROC plots, sensitivity (d′) and response bias (c) were calculated for each SOA and plotted for the flashbeep (B, C) and speech stimuli (E, F), respectively. Error bars indicate 95% confidence intervals of the mean. * = non-overlapping confidence intervals

For the AV TOJ flashbeep task (Fig. 5.5a), the ROC plot reveals that there are nearly equivalent shifts in sensitivity between groups (Fig. 5.5b). Thus, there are no differences between groups for d′, although the lower overall sensitivity for speech TOJ (Fig. 5.5e) compared with flashbeep TOJ (Fig. 5.5b) illustrates the comparatively greater difficulty of the speech task (i.e., maximum d′ of 1.5 v. 2.4). Furthermore, bias appears comparable between the two groups, except for one significant difference at the 50 ms SOA (Fig. 5.5c). In contrast, for the AV TOJ speech task, there are apparent differences in the ROC plot (Fig. 5.5d) such that CI users have more 'visual first responses.' Again, there are no differences in sensitivity between groups (Fig. 5.5e); however, response bias differs significantly between groups for all but the two shortest SOAs (50 and 100 ms). This finding reflects a visual bias for CI users toward a greater likelihood of selecting the 'visual first' response (Fig. 5.5f). In summary, data from these AV TOJ tasks reveal strikingly similar performance between groups for the flashbeep task, similar sensitivity for the speech task, and substantially different response biases for the speech task.

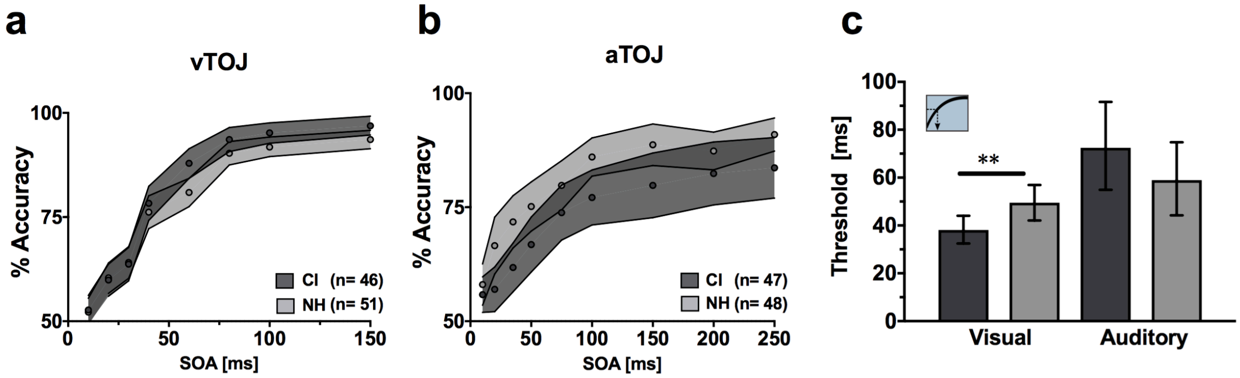


FIGURE 5.6. Unisensory temporal order judgment tasks. Mean accuracies per SOA are plotted for visual (A) and auditory (B) TOJ tasks. CI users are shown in dark gray and NH in light gray. Circles indicate means and shaded areas correspond to 95% confidence intervals. Threshold calculations (C) at 75% accuracy are shown for both tasks with error bars also indicating 95% confidence intervals. CI = cochlear implant; NH = normal hearing; SOA = stimulus onset asynchrony; aTOJ = auditory temporal order judgment; vTOJ = visual temporal order judgment; $^{**}p < 0.01$

Visual TOJ thresholds are improved in CI users, but auditory thresholds do not significantly differ from controls. To illustrate temporal unisensory performance across groups, we plotted performance at each tested SOA for the CI users and NH controls. Group averages for visual TOJ (A) and auditory TOJ (B) are plotted in Figure 5.6. As noted previously (Table 5.2), thresholds measured at 75% accuracy were significantly different between groups for the visual TOJ task but not the auditory TOJ task (Fig. 5.6c). A post hoc MANCOVA for the vTOJ task (dependent variables: 8 SOAs, covariate: age) did not indicate any further statistically significant group differences ($F_{(8,88)}$= 1.74, p= 0.1; Wilk's Λ = 0.863, η_p^2= 0.137). Therefore, significant group differences in vTOJ are limited to threshold measures (Fig. 5.6c) and not any broader differences across individual SOAs (Fig. 5.6a). Because a difference in threshold could result from either differences in low level sensory processing (i.e., sensitivity or discriminability changes) or higher level decisional factors (i.e., bias or criterion changes), we further investigated these results using SDT.

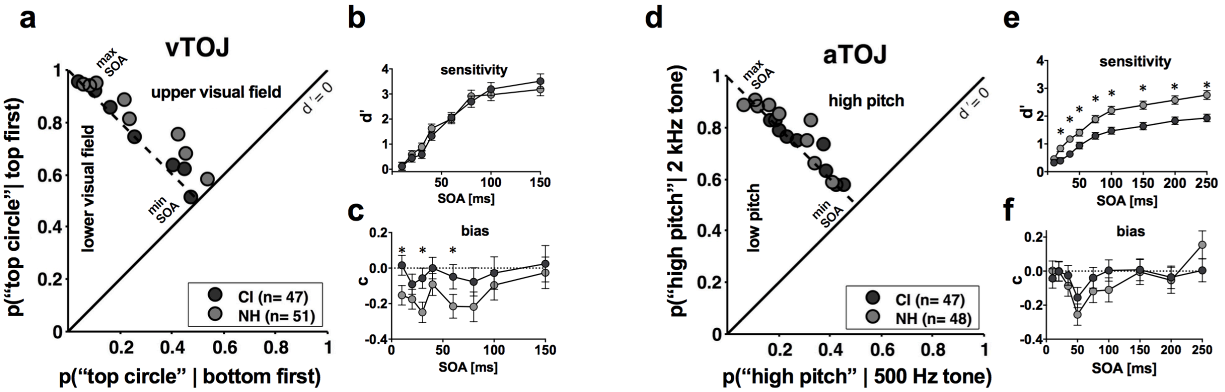


FIGURE 5.7. Signal detection analysis of unisensory temporal order judgment tasks. The probability of hits versus probability of false alarms are plotted for the vTOJ (A) and aTOJ tasks (D). From these ROC plots, sensitivity (d') and response bias (c) were calculated for each SOA and plotted for vTOJ (B, C) and aTOJ (E, F), respectively. Error bars indicate 95% confidence intervals of the mean. * = non-overlapping confidence intervals

Signal detection analysis reveals differences in aTOJ sensitivity and vTOJ response bias. The probability of correct responses to negative SOAs is plotted on the y axis versus incorrect responses to positive SOAs on the x axis in ROC space for both the visual (Fig. 5.7a) and auditory TOJ tasks (Fig. 5.7d). For the visual task, negative SOAs were arbitrarily defined as trials where the top circle appeared first, and for the auditory task, when the high pitch was presented first. For the vTOJ task we see equivalent sensitivity (Fig. 5.7b) yet a shift in response bias such that CI users are more like unbiased observers compared to NH controls. Curiously, this difference in the NH group manifests as a bias in reporting the stimulus in the upper visual field appearing first (Fig. 5.7c). For the aTOJ task (Fig. 5.7d), sensitivity is lower in the CI group at all but the most difficult SOA (Fig. 5.6e). In contrast, response bias is unaffected in the aTOJ task (Fig. 5.7f). Together, these findings suggest that: 1) CI users employ distinct response strategies in the vTOJ task that minimize response biases, and 2) CI users have decreased sensitivity for the aTOJ task.

5.4 Discussion

A key finding in this study is a shift in the point of subjective simultaneity (PSS) for making temporal judgments regarding audiovisual speech in postlingually deafened adults with CIs compared to NH controls. Specifically, the NH control group required visual speech cues to precede auditory speech cues by 39 ms more than CI users at the point of maximum perceived simultaneity in the SJ task (Fig. 5.3c). This finding is consistent with our expectation of a PSS more centered near objective synchrony; however, we expected this shift to be accompanied by a broadening of the overall width of the temporal binding window for speech stimuli (Fig. 5.8a), a result that was not supported by the data. Rather, CI users had lower reports of simultaneity at auditory-leading speech SOAs (Fig. 5.3b), resulting in comparable TBWs for both groups (Fig. 5.8b). Thus, compared to NH controls, CI users have greater accuracy identifying temporal asynchronies with visual-leading speech, yet lower accuracy with auditory-leading speech.

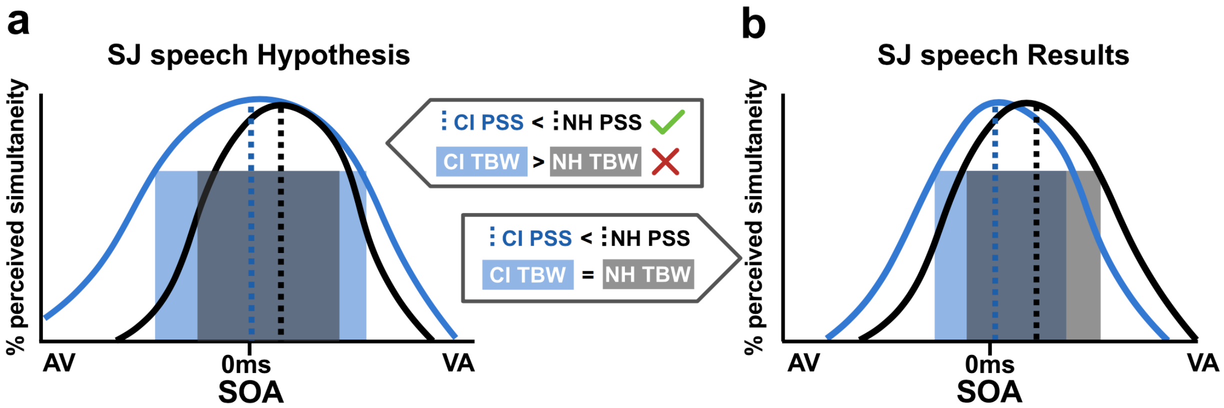


FIGURE 5.8. Summary of the main hypothesis and results. Although the direction of our hypothesized speech PSS shift (A) was supported by our results (B), the speech TBW for CI users was non-significantly different from controls and not extended as we had anticipated.

Interestingly, from the only other published test of simultaneity judgments in CI users, Hay-McCutcheon et al. report a similar, albeit statistically non-significant trend of CI users in a similar age group having a smaller PSS (i.e., CI = 58 ms; NH = 72 ms). It is possible that their small sample (n = 12) contributed to a statistically underpowered comparison in the self-described preliminary analysis (Hay-McCutcheon et al., 2009). If so, the same leftward shift may exist for both syllables (shown here) as well as full words. To our knowledge, the significant speech PSS shift reported here is a novel finding and one that warrants further investigation to more conclusively establish whether broader generalization exists with other speech cues (e.g., full words and sentences).

We first questioned how any artificial auditory latency introduced by CI processors may have influenced our group-level differences in speech PSS. Although exact processing times vary by the manufacturer, sound processor, and programmed pulse rate, the magnitude is generally on the order of 10 ms. Because this processing is bypassing acoustic conduction in the middle and inner ear, actual first spike latencies in the VIII cranial nerve in CI users compared to NH controls would actually differ by less than 10 ms. As a result, SOAs in our study may have an artificial auditory delay of at most 10 ms in the CI group. If that is the case, then SOAs for CI users are actually *more* visually-leading, which is rightward and in the opposite direction of our findings of *less* visually-leading subjective synchrony in CI users (i.e., leftward shifts closer to objective synchrony at 0 ms). Therefore, if greater auditory latency was introduced by CI processors, then the group-level differences in SJ speech PSS are even more robust.

The observed shift in PSS in the CI users could be a result of changes in so-called “bottom up” (i.e., stimulus-dependent) factors or due to changes in more “top-down” (i.e., decisional) factors such as response bias. To delve further into these differences, we utilized a different temporal paradigm – the temporal order judgment task – that allows for a dissection of these factors using principles derived from signal detection theory (SDT) and, specifically, the calculation of measures of sensitivity and response bias. Although not strictly adhering to SDT, these analyses strongly suggest no differences between groups in AV temporal sensitivity (for either flashbeep or speech stimuli), but a significant difference in response bias for the speech stimuli, manifesting as a strong visual bias for the CI users (Fig. 5.5f).

We then considered whether, in the absence of any prompting from the task instructions, CI users had preferentially directed their attention toward the visual speech component throughout the SJ task. The effects of such attentional cueing have been measured in a number of studies investigating attention-dependent sensory acceleration, also known as “prior entry” (Schneider and Bavelier, 2003; Spence and Parise, 2010; Zampini et al., 2005). For example, endogenous cueing that overtly directs a viewer’s attention toward the visual component in an SJ task causes the PSS to shift leftward by 14 ms when stimuli are short noise bursts and illuminated LEDs (Zampini et al., 2005). Given the higher complexity of speech syllables as well as the temporal ambiguity introduced from articulation onset to voicing onset, we consider a 39 ms SJ PSS shift here as a reasonable magnitude to fall within this explanation. Interestingly, physically manipulating a visual stimulus to be more salient than an auditory cue (i.e., exogenous cueing) also shifts PSS in a similar manner (Boenke et al., 2009). Broadly speaking, highly salient, attention-grabbing stimuli of various types (i.e., either crossmodal or intramodal) are well-known to be perceived as occurring prior to less salient ones (Krekelberg and Lappe, 2001; Moutoussis and Zeki, 1997; Zampini et al., 2005). Thus, in the absence of any overt attentional cueing in the instructions or by the researchers, CI users’ SJ speech curves appear biased toward vision (fewer reports of synchrony at +SOAs) at the expense of auditory-leading trials (more reports of synchrony at -SOAs). Therefore, an attend-vision response strategy in the CI group may explain an overall leftward shift in PSS as seen in the speech SJ task without a constriction of the TBW (Fig. 5.8b).

This finding is novel in that it suggests that CI users are employing greater visual weighting in temporal judgments of speech (when compared with NH individuals). Thus, for CI users, low reliability of an auditory sensory estimate likely results in placing lower weight on the auditory information in the process of AV cue combination (van Dam et al., 2014). It seems plausible that daily, focused lip reading while listening with a CI causes higher perceptual weighting of the visual speech signal. Interestingly, a recent study in pediatric CI users reported lower auditory dominance for temporal judgments, and this lessened auditory weighting had a negative impact on language skills in a prelingually deafened cohort (Gori et al., 2017).

Unlike the paucity of data surrounding AV temporal processing in CI users, perception of AV syllables has been extensively studied via a common proxy for multisensory integration called the McGurk effect (McGurk and MacDonald, 1976; Stevenson et al., 2017a). In this crossmodal illusion, conflicting AV syllables can elicit a novel percept. For example, a sound file of the syllable "ba" dubbed onto the visual articulation of "ga" often elicits the perception of a third syllable such as "da" or "tha" for the viewer. Presentation of these incongruent stimulus pairings creates a perceptual discrepancy that drives individuals to report either the fused multisensory percept or the token for the sense providing the best sensory estimate (i.e., visual or auditory capture). Highly consistent results across several studies of the McGurk illusion indicate a visual bias in CI users that is rarely seen in NH individuals (Desai et al., 2008b; Huyse et al., 2013; Rouger et al., 2008; Schorr et al., 2005; Tremblay et al., 2010). Our results here using the same syllables from the McGurk illusion ("ba" and "ga") suggests that temporal judgements of these AV cues are also visually-biased (Ipser et al., 2017).

Turning to the unisensory TOJ tasks, we also found differences in vTOJ thresholds (Fig. 5.6c) and response bias between groups (Fig. 5.7c). A possible explanation for these differences is that the task required subjects to fixate on a cross in the center of the screen and distribute their spatial visual attention to monitor two locations in the upper and lower visual fields. It is possible that the ability of CI users to perform more like unbiased observers reflects enhanced attentional allocation to the relevant parafoveal visual locations (Parasnis and Samar, 1985). NH controls were seemingly less able to do this, and instead focused more on the upper visual field. In the NH group, this response bias may have resulted for several plausible reasons. One possibility is that visual apparent motion may have been experienced as a result of rapidly flashing the circles in quick succession. In the absence of well-distributed visual attention, NH controls may simply have responded according to known anisotropies in visual apparent motion detection to favor downward perceived motion or “top-first” responses. Such visual motion biases are frequently reported yet are highly dependent on specific task parameters (Skottun and Skoyles, 2010). Thus, without further testing, it is difficult for us to conclude to what extent this played a role in NH response bias.

Interestingly, reduced response bias in the CI group also corresponded to group differences in vTOJ thresholds. Accuracy in this task has previously been shown to significantly decrease with age in typical individuals (Stevenson et al., 2017b). Not surprisingly, group differences were only evident when age was included as a covariate in the comparison to correct for the fact that the CI group was older by 8.4 years. A preliminary study from our group also suggests that visual temporal thresholds in prelingually deafened adults with CIs are predictive of speech comprehension (Jahn et al., 2017). In the future, closer age-matching between groups, particularly with the vTOJ task, will better eliminate this potential confound and allow us to further investigate between-group differences.

Although aTOJ was equivalent for CI users and controls at threshold (Table 5.2), CI users seemed to exhibit lower performance across most SOAs (Fig. 5.6b) that was not evident in the global threshold measurements (Fig. 5.6c). Reduced accuracy as well as d’ across many SOAs (Fig. 5.7e) suggests that the frequency discrimination in this task may be comparatively more difficult for CI users. That is, lower performance even at the largest SOA does not suggest an auditory temporal processing deficit per se but rather a confounding factor of frequency discrimination inherent to the task (for a similar discussion see (Skottun and Skoyles, 2010). Although 500 Hz and 2 kHz tones were detectable for CI users, we believe that the necessary discrimination overshadowed our ability to measure temporal processing by itself. In fact, this frequency component may also explain why we have consistently found auditory thresholds to be larger than visual thresholds with this task (Fig. 5.6c) (Stevenson et al., 2014c, 2017b, 2017c), despite the auditory system’s specialization for temporal processing. Although it is inconclusive on the basis of this task whether auditory temporal processing was intact in our CI cohort, several prior studies of gap detection thresholds indicate normal thresholds in CI users, particularly those with postlingual onset of deafness (Busby and Clark, 1999; Shannon, 1989). Furthermore, gap detection performance is known to approach adult-like thresholds by adolescence (Irwin et al., 1985), which in our cohort, would have occurred prior to severe-to-profound hearing loss. Thus, while we do not anticipate any auditory temporal impairments in the postlingually deafened CI users studied here, we cannot rule out this possibility.

A notable limitation to our interpretations of all signal detection analyses is that without having counterbalanced responses to different numbers, biases toward vision and audition, for instance, cannot be distinguished from preference for the numbers 1 or 2 (see instructions in Fig. 5.1). This could be a considerable confound when testing children; however, in adults who confirmed understanding of the task, such superficial biases seem less likely. Furthermore, future incorporation of reaction time measures into these tasks may further reveal perceptual differences with speeded judgments as others have shown in deafness (Nava et al., 2008).

In conclusion, we show here that adult CI users judge the temporal relationship between auditory and visual speech in a visually-biased manner; however, the benefits (or consequences) of this weighting remain unknown. Ongoing work in our laboratory aims to elucidate how the present findings for temporal processing differences map onto the considerable clinical diversity in this population and has the potential to yield important insights—for example, into how the aforementioned results relate to AV integration of words, auditory-only speech recognition, and perhaps even broader language comprehension in CI users. Future investigations exploring the impact and generalization of temporal training (De Niear et al., 2018; Fujisaki et al., 2004; Powers et al., 2009; Powers III et al., 2016) will additionally be beneficial for addressing whether remediation of altered temporal perception can positively impact AV gain for CI users. Ideally, this intervention could afford users with the maximum possible benefit from their CIs, which is closely tied with quality of life measures in this rapidly growing clinical population.

5.5 Acknowledgments

This work has been published in the journal Scientific Reports:

Butera, I.M., Stevenson, R.A., Mangus, B.D., Woynaroski, T.G., Gifford, R.H., and Wallace, M.T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. Sci. Rep. *8*, 11345.

CHAPTER 6 A summary of findings and conclusions

6.1 Summary of findings

The table below summarizes the key findings from each of these studies.

CH	Title	Key findings
1	An introduction to speech perception with a cochlear implant	• CI outcome measures are a) highly variable and b) are primarily assayed in auditory-only listening conditions • Visual plasticity during deafness may influence multisensory processing postoperatively • Audiovisual speech and its neural correlates may provide further insight into individual differences in how patients utilize CI-mediated hearing for speech understanding
2	Visually-biased illusion perception in CI users	• CI users have lower AV illusion perception with both speech stimuli (McGurk) and simple, flashbeep stimuli (SIFI) • Duration of hearing impairment is a significant predictor of CNC scores, but neither McGurk nor SIFI results explained additional CNC variability
3	Functional localization of AV speech using fNIRS	• Our behavioral and fNIRS paradigms are able to quantify significant AV gain in a cohort of normal hearing listeners • McGurk perception does not correlate to semantic AV word recognition tasks, though it does have a positive correlation to visually-evoked activity during lipreading
4	Behavioral and functional neuroimaging of CI users' AV processing	• Fewer CI users perceive a fused McGurk syllable. Instead, most report the visual syllable /ga/. • Though CI users are better lipreaders, they experience the same overall magnitude of AV gain as NH controls • High performing CI users are less proficient lipreaders and have more auditory-evoked cortical activity in the left temporal lobe. • The McGurk task does not correlate with other AV tasks
5	AV temporal processing in postlingually deafened adults with CIs	• Speech PSS is shifted closer to objective synchrony in CI users, while flashbeep PSS and TBWs are unaffected • Visual-only temporal order judgments (TOJ) are enhanced in CI users (i.e., CI thresholds are lower) • CI users have a visual response bias in an AV TOJ task with speech, while flashbeep TOJ is unaffected

Table 6.1. Overview of key findings by chapter. AV = audiovisual, fNIRS= functional near infrared spectroscopy, CI = cochlear implant, NH = normal hearing, PSS = Point of objective synchrony, TBW = temporal binding window, TOJ = temporal order judgement

6.2 A unifying theme of "visuocentric" listening in cochlear implant users

The underlying theoretical framework for this dissertation is based upon the known degradation of spectrally-complex stimuli for CI users and the significant challenge that this poses for the brain. Specifically, degradation of incoming auditory stimuli should, in theory, render the CI recipient more dependent upon visual information for successful communication (Massaro, 1987; Massaro and Cohen, 2000; Massaro et al., 1988). Because auditory features transmitted by a CI are spectrally degraded, they would be modeled as having a low weight. As such, the visual features should, in theory, hold greater weight. The preferential weighting of visual information over auditory information is a common theme throughout the results of these experiments, which could be summarized as a natural tendency by CI users to engage in a type of "visuocentric" listening.

6.3 Consistency in McGurk findings

With more than double the sample size of any prior study, chapter 2 and 4 represent the largest assay to-date of the McGurk illusion in CI users. For a brief overview of these results, all data is plotted in Figure 6.1 for each chapter (a,b) and of all individuals combined (c,d). Although a decrease in p(McGurk) for CI users was only significant in chapter 2 after excluding non-perceivers, the results in chapter 4 alone and of all these data combined (Fig. 6.1c), support the conclusion that this illusion is experienced more often by those with normal hearing. Though /da/ responses are roughly similar between groups (Fig. 6.1d), many of these responses are likely to be mistakenly-identified auditory percepts in the CI group and not McGurk percepts per se. This important issue is taken into account in the derived p(McGurk) measure, and it is possible that a reanalysis of prior studies may result in new interpretations more aligned with what we report here.

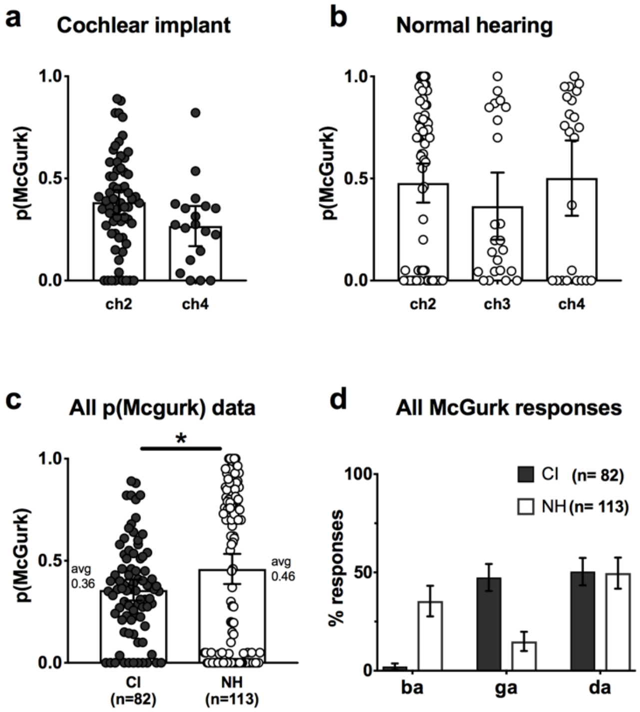


FIGURE 6.1. McGurk data by chapter and compiled.

6.4 Comparing fNIRS results between chapters 3 and 4 for NH listeners

There are several differences in the fNIRS findings in chapters 3 and 4. Although the tasks and analysis are very similar, the fNIRS equipment and data collection methods are quite different. The NIRx equipment used in chapter 3 had many more optodes that were fit into a 10-20 cap (Fig. 6.2a). The 52 channels that resulted allowed us to investigate a much broader cortical area (Fig. 6.2b).

Activity measured in the left auditory channel, containing primary and associative cortices, was significant on both hemispheres (Fig. 6.2c). Interestingly, in comparing the magnitude of this change in oxyHb and deoxyHb concentration, it is actually quite small in comparison to chapter 4 data. This is immediately apparent when the y axis is scaled the same as the data collected from the Hitachi equipment (Fig. 6.2d). The use of lasers instead of LEDs may play a role on account of the higher intensity of lasers, and greater capacity to record a signal change (Fig. 6.2g).

It's also possible that the different methods of aligning probes with 10-20 points contributed to larger confidence intervals in the Hitachi data (Fig. 6.2g). In this setup (Fig. 6.2e), the two halves of the holders move independently and may have less consistent alignment between individuals, which could negatively impact the group averages of activity using this technique. This may have been an issue, for example, in the data for the left auditory areas represented by channel 6, which did not have significant activity (Fig. 6.2g). This result is in sharp contrast to the strong activity and typical time course observed in the same probe in chapter 3 (i.e., channel 10; Fig. 6.2c).

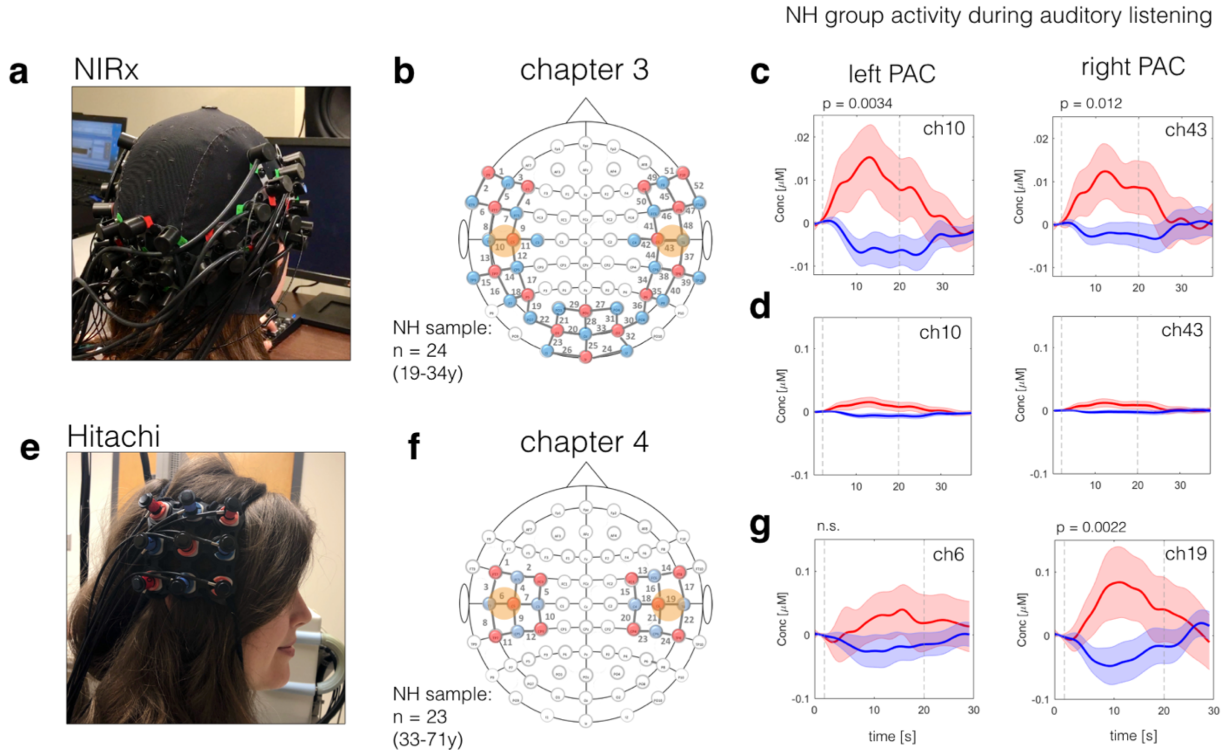


FIGURE 6.2. A comparison of NH fNIRS data between two types of equipment. Data from the NIRx equipment in chapter 3 (a,b) had a high signal-to-noise ratio (c), yet a small overall magnitude (d) in comparison to the Hitachi equipment used in chapter 4 (e-g). NH block averages for auditory-evoked activity lie within the orange circles (in b,f). Red = oxyHb, Blue = deoxyHb, shaded areas are 95% confidence intervals of the mean.

Importantly, these two types of NIRS equipment emit different wavelengths of light as well. For deoxyHb and oxyHb respectively, the wavelength pairs are 760/850nm with Hitachi and 695/830nm with NIRx equipment. Given the known differences in SNR depending on wavelength combinations, this discrepancy could have been a between-studies factor as well (Uludağ et al., 2004).

There are also differences in the age of subjects between the two studies. NH controls in chapter 3 range from 19-24 years of age, while chapter 4 participants are quite a bit older (33-71y). We used the same source-detector spacing of 2 cm for both studies. If there were any substantial differences in dura, skull, or skin thickness with age, then it's possible that this distance should be adjusted (Quaresima et al., 2012).

In future studies, a 3D digitizer would also improve the registration of individual optodes and potentially the replicability of results. Furthermore, the addition of short channel recordings could enable regressing out the extracerebral signal changes, which is a technique that's rapidly gaining popularity in fNIRS studies (Goodwin et al., 2014). In lieu of bandpass filtering and motion correction, this analytic approach enters systemic changes into a glm to better identify functional brain activity and thereby, increase the likelihood of detecting more subtle effects. Along the same lines, more dense recording with many overlapping channel distances can allow for a more detailed understanding of activity in a small area of interest (Hassanpour et al., 2015; Olds et al., 2016; Pollonini et al., 2014).

The visually-evoked activity identified in chapter 3 during lipreading would be particularly interesting to investigate in CI users. The rigid 3x3x2 optode arrangement in our study with cochlear implant users (Fig. 6.2e) is not well suited for imaging this area. Given the overlap between the region of interest and implants, it may only be possible to compare visual motion processing in unilateral CI users with one unobstructed hemisphere.

6.5 Future clinical applications for predicting outcomes from variations in neurophysiology and behavior

Receiving a cochlear implant is an irreversible surgical procedure and managing patients' expectations is a major clinical concern. In order for patients to make informed decisions, clinicians need to have reasonable predictions of performance to effectively counsel patients about the postoperative speech understanding ability they could expect to achieve. Although neuroimaging studies have enabled quite a bit of understanding of brain plasticity during deafness and after cochlear implantation, only recently is this work being applied to predictive efforts. For instance, structural MRIs of early-deafened kids below the age of 3.5 y were used to build machine learning models to predict post-op speech performance (Feng et al., 2018). Another recent study used a functional MR approach to build a model of language-learning outcomes based on a listening task in the scanner (Tan et al., 2015). The incorporation of structural, functional, or other physiological metrics to indicate the aptitude of an individual's central and peripheral nervous system for utilizing a CI-mediated sense of sound would enable patients and providers to make more informed decisions. These include when to undergo CI surgery, in which ear, at what age, or whether to have the surgery at all. Many patients with age-related hearing loss are even opting for CIs in the last years of life. These and all other patients and caregivers considering CI surgery could benefit from better judging what post-implantation hearing may be like for them. For all the benefits that cochlear implants can provide, simply having one of these neuroprosthetics is insufficient for full hearing restoration. Fortunately, vision can help.

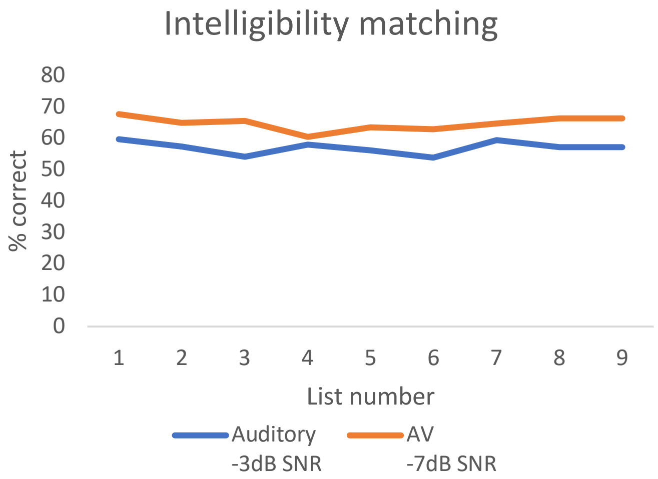
APPENDIX

A. fNIRS word lists. Testing included 4 categories of 20 words each for a total of 80 words.

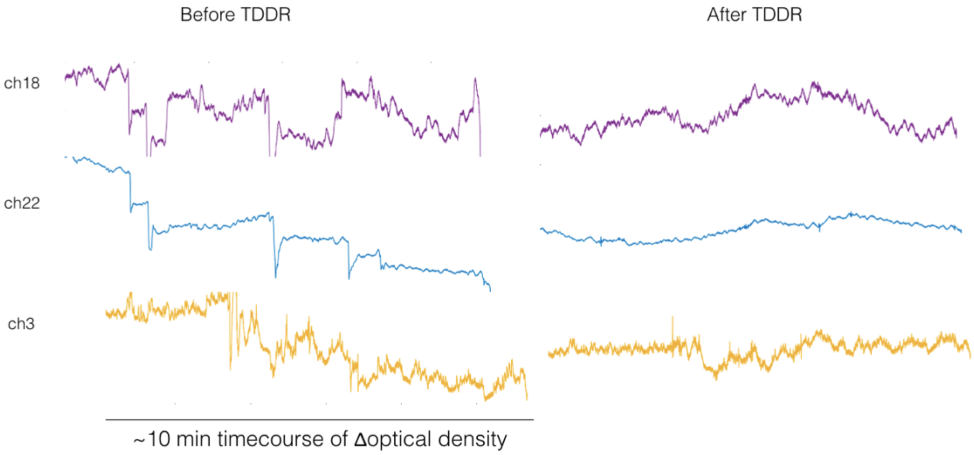
	fNIRS Run 1		fNIRS Run 2	
	Objects	Numbers	Animals	Actions
1	ball	one	bear	cheer
2	bed	two	bee	read
3	bell	three	bird	knock
4	bike	four	bug	thank
5	book	five	calf	tell
6	box	six	cat	wish
7	chair	seven	crab	chat
8	clock	eight	dog	ride
9	cup	nine	duck	run
10	map	ten	fish	walk
11	nail	eleven	frog	toss
12	note	twelve	goose	look
13	page	thirteen	hen	fold
14	rag	fourteen	horse	catch
15	ring	fifteen	moth	press
16	shirt	sixteen	mouse	guess
17	shoe	seventeen	rat	teach
18	spoon	eighteen	sheep	fall
19	truck	nineteen	snake	take
20	wheel	twenty	toad	cut

	A High SNR	AV High SNR	V High SNR	A Med SNR	AV Med SNR	V Med SNR	A Low SNR	AV Low SNR	V Low SNR
	List 1	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**
1	No	Dodge	Shoe	Tar	Sung	Fail	Yam	Which	Match
2	Look	Nail	Tar	Quick	Raw	Learn	Chat	Snake	Sane
3	Jug	Tray	Leave	Raid	Loose	Yam	Choose	Kite	Germ
4	Beg	Week	Gone	Road	Paste	Life	If	Pinch	Web
5	Shore	Kite	Beet	Cheek	Hush	Patch	Big	Hiss	Soon
6	Head	Judge	Hall	Neat	Lack	Have	Rug	Kick	Fudge
7	Meek	Pool	Run	Vine	Haze	Near	Limb	Coat	Mouth
8	Shawl	Soap	Knock	Thank	For	Burn	Wife	Peak	Rob
9	Bought	Low	Thank	Doll	Note	Shout	Kick	Ship	Map
10	Class	Guess	Get	Foot	Hike	Few	Rich	Path	Rough
11	Long	Wish	Tooth	Cool	Is	Freeze	Bush	Your	Pole
12	Lean	Chill	Pearl	Bud	Gray	Them	Hole	His	Fan
13	Tip	Cool	Phone	Puff	Wash	Base	Camp	Sung	Sour
14	Tell	Fail	Weight	Sob	Tongue	Cut	Sin	Slice	Cab
15	Nuts	Fit	Juice	Veal	Loud	Beet	Tape	Chore	Led
16	Beef	Raid	Loaf	Rat	Mop	Pearl	Girl	Wrong	Gum
17	Birth	Press	Merge	Dig	Rough	Purge	Scab	Clock	Rib
18	Boat	Cab	Sob	Shock	Pod	Tray	Death	White	Bone
19	Sun	What	Sour	Fit	Geese	Bean	Path	Set	Nice
20	Weed	Third	Clown	Gap	Keep	Falls	Sure	Third	Cause
21	Nail	Those	Hand	Deck	Mouse	Food	Read	Long	End
22	Dish	Calm	Led	Get	Jay	Doom	Gas	Loop	Such
23	Noise	Vote	Came	Hike	As	Jade	Else	Make	Late
24	Cage	Wheel	Own	Love	Said	On	Nice	Need	Roof
25	Cape	Youth	Laugh	Peg	Red	Thick	Me	Great	Wag
26	Mess	Void	Half	Page	Search	Loose	Grab	Doom	Peace
27	Goose	Ton	Home	Mine	Which	Judge	Rat	Move	Watch
28	Mill	Niece	Purge	Mode	Shack	Nick	Shore	Grew	Sin
29	Shine	Goal	Set	Weight	Coin	Moon	Meek	Chalk	Perch
30	Loop	Live (ai)	Most	Tool	Gin	Sag	Nap	Time	Safe
31	School	Fade	Tick	Raw	Yearn	Bee	Bone	Kid	Cup
32	Fair	Are	End	Live	Fine	See	Axe	Chief	Team
33	Sail	White	Slice	Meet	Slip	Move	Dead	Case	Feet
34	Whip	Life	Sink	Chin	Pants	As	Gale	Blind	Hull
35	That	Knife	Rain	Wish	Mode	Mood	Ripe	Burn	Teach
36	Black	Lip	Hit	Train	Jar	Next	Wide	Hair	Press
37	Hole	Mate	Reach	Rig	Mood	Note	Tire	Ball	Sink
38	House	Take	Kill	Mop	Deep	Half	Pain	Crab	Nest
39	Name	Air	Gum	Gore	Hull	Sure	Check	Sack	Hunt
40	Low	Tool	Jade	Rot	Drop	Cake	South	Next	Waste

B. Word recognition word lists. Testing included 9 word lists with 286 unique words and 360 trials total.



C. Intelligibility matching word lists. The 9 word lists for the word recognition testing was matched for intelligibility based on word identification in -3 and -7 dB SNRs by 20 normal-hearing participants as described by Picou et al. (2011).



D. Example fNIRS filtering with Temporal Derivative Distribution Repair (TDDR). Three example channels containing motion artifacts were corrected.

REFERENCES

Anderson, C.A., Wiggins, I.M., Kitterick, P.T., and Hartley, D.E.H. (2017). Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. Proc. Natl. Acad. Sci. *114*, 10256–10261.

Barone, P., and Deguine, O. (2011). Multisensory Processing in Cochlear Implant Listeners. In Auditory Prostheses, F.-G. Zeng, A.N. Popper, and R.R. Fay, eds. (New York, NY: Springer New York), pp. 365–381.

Barone, P., Strelnikov, K., and Deguine, O. (2010). Auditory recovery and speechreading in cochlear implanted deaf patients: A review. Audiol. Med. *8*, 100–106.

Baskent, D., and Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. Ear Hear. *32*, 582–592.

Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Corina, D., Liu, G., and Neville, H. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. J. Neurosci. *20*, RC93.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., and Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. J. Neurosci. *21*, 8931–8942.

Bavelier, D., Dye, M.W.G., and Hauser, P.C. (2006). Do deaf individuals see better? Trends Cogn. Sci. *10*, 512–518.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., and Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. *7*, 1190–1192.

Beauchamp, M.S., Nath, A.R., and Pasalar, S. (2010). fMRI-Guided Transcranial Magnetic Stimulation Reveals That the Superior Temporal Sulcus Is a Cortical Locus of the McGurk Effect. J. Neurosci. *30*, 2414–2417.

Bergeson, T.R., Pisoni, D.B., and Davis, R.A.O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear Hear. *26*, 149–164.

Bilger, R.C., and Hopkinson, N.T. (1977). Hearing Performance with the Auditory Prosthesis. Ann. Otol. Rhinol. Laryngol. *86*, 76–76.

Bisconti, S., Shulkin, M., Hu, X., Basura, G.J., Kileny, P.R., and Kovelman, I. (2016). Functional near-infrared spectroscopy brain imaging investigation of phonological awareness and passage comprehension abilities in adult recipients of cochlear implants. J. Speech Lang. Hear. Res. *59*, 239–253.

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindström, B., Nedzelski, J., Peterson, A., et al. (1996). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants. Audiol. Neurotol. *1*, 293–306.

Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., et al. (2013). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients. Audiol. Neurotol. *18*, 36–47.

Boenke, L.T., Deliano, M., and Ohl, F.W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. Exp. Brain Res. *198*, 233–244.

Bortfeld, H. (2019). Functional near-infrared spectroscopy as a tool for assessing speech and spoken language processing in pediatric and adult cochlear implant users. Dev. Psychobiol. *61*, 430–443.

Bosworth, R.G., and Dobkins, K.R. (1999). Left-hemisphere dominance for motion processing in deaf signers. Psychol. Sci. *10*, 256–262.

Bosworth, R.G., and Dobkins, K.R. (2002). The Effects of Spatial Attention on Motion Processing in Deaf Signers, Hearing Signers, and Hearing Nonsigners. Brain Cogn. *49*, 152–169.

Bottari, D., Nava, E., Ley, P., and Pavani, F. (2010). Enhanced reactivity to visual stimuli in deaf individuals. Restor. Neurol. Neurosci. *28*, 167–179.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 433–436.

Burr, D., and Alais, D. (2006). Chapter 14 Combining visual and auditory information. In Progress in Brain Research, Martinez-Conde, Macknik, and Martinez, Alonso & Tse, eds. (Elsevier), pp. 243–258.

Busby, P.A., and Clark, G.M. (1999). Gap detection by early-deafened cochlear-implant subjects. J. Acoust. Soc. Am. *105*, 1841–1852.

Butera, I.M., Stevenson, R.A., Mangus, B.D., Woynaroski, T.G., Gifford, R.H., and Wallace, M.T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. Sci. Rep. *8*, 11345.

Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C., and Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport *14*, 2213–2218.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science *276*, 593–596.

Cecere, R., Rees, G., and Romei, V. (2015). Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. Curr. Biol. *25*, 231–235.

Chen, Q., Zhang, M., and Zhou, X. (2006). Effects of spatial distribution of attention during inhibition of return (IOR) on flanker interference in hearing and congenitally deaf people. Brain Res. *1109*, 117–127.

Ching, T.Y.C., Day, J., Buynder, P.V., Hou, S., Zhang, V., Seeto, M., Burns, L., and Flynn, C. (2014). Language and speech perception of young children with bimodal fitting or bilateral cochlear implants. Cochlear Implants Int. *15*, S43–S46.

Collignon, O., Champoux, F., Voss, P., and Lepore, F. (2011). Chapter 14 - Sensory rehabilitation in the plastic brain. In Progress in Brain Research, A.M. Green, C.E. Chapman, J.F. Kalaska, and F. Lepore, eds. (Elsevier), pp. 211–231.

van Dam, L.C.J., Parise, C.V., and Ernst, M.O. (2014). Modeling Multisensory Integration. In Sensory Integration and the Unity of Consciousness, D.J. Bennett, and C.S. Hill, eds. (The MIT Press), pp. 209–229.

De Niear, M.A., Gupta, P.B., Baum, S.H., and Wallace, M.T. (2018). Perceptual training enhances temporal acuity for multisensory speech. Neurobiol. Learn. Mem. *147*, 9–17.

Deggouj, N., Gersdorff, M., Garin, P., and Castelein, S. (2007). Today's indications for cochlear implantation. Todays Indic. Cochlear Implant.

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

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

Donley, J., Ritz, C., and Kleijn, W.B. (2018). Multizone Soundfield Reproduction With Privacy- and Quality-Based Speech Masking Filters. IEEEACM Trans. Audio Speech Lang. Process. *26*, 1041–1055.

Donley, Jacob (2017). Sound Zone Tools (Github repository).

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

Duchesne, L., Millette, I., Bhérer, M., and Gobeil, S. (2017). Auditory performance and subjective benefits in adults with congenital or prelinguistic deafness who receive cochlear implants during adulthood. Cochlear Implants Int. *18*, 143–152.

Dye, M.W.G., Hauser, P.C., and Bavelier, D. (2009). Is Visual Selective Attention in Deaf Individuals Enhanced or Deficient? The Case of the Useful Field of View. PLOS ONE *4*, e5640.

Eijk, R.L.J. van, Kohlrausch, A., Juola, J.F., and Par, S. van de (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. Percept. Psychophys. *70*, 955–968.

Erickson, L.C., Zielinski, B.A., Zielinski, J.E.V., Liu, G., Turkeltaub, P.E., Leaver, A.M., and Rauschecker, J.P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. Front. Psychol. *5*.

Ernst, M.O., and Bülthoff, H.H. (2004). Merging the senses into a robust percept. Trends Cogn. Sci. *8*, 162–169.

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical Evidence of Multimodal Integration in Primate Striate Cortex. J. Neurosci. *22*, 5749–5759.

Feng, G., Ingvalson, E.M., Grieco-Calub, T.M., Roberts, M.Y., Ryan, M.E., Birmingham, P., Burrowes, D., Young, N.M., and Wong, P.C.M. (2018). Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients. Proc. Natl. Acad. Sci. *115*, E1022–E1031.

Ferradal, S.L., Eggebrecht, A.T., Hassanpour, M., Snyder, A.Z., and Culver, J.P. (2014). Atlas-based head modeling and spatial normalization for high-density diffuse optical tomography: In vivo validation against fMRI. NeuroImage *85*, 117–126.

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. NeuroImage *63*, 921–935.

Fine, I., Finney, E.M., Boynton, G.M., and Dobkins, K.R. (2005). Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. J. Cogn. Neurosci. *17*, 1621–1637.

Finley, C.C., Holden, T.A., Holden, L.K., Whiting, B.R., Chole, R.A., Neely, G.J., Hullar, T.E., and Skinner, M.W. (2008). Role of Electrode Placement as a Contributor to Variability in Cochlear Implant Outcomes: Otol. Neurotol. *29*, 920–928.

Finney, E.M., Fine, I., and Dobkins, K.R. (2001). Visual stimuli activate auditory cortex in the deaf. Nat. Neurosci. *4*, 1171–1173.

Fishburn, F.A., Ludlum, R.S., Vaidya, C.J., and Medvedev, A.V. (2019). Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. NeuroImage *184*, 171–179.

Fitzpatrick, E. (2015). Chapter 19 - Neurocognitive development in congenitally deaf children. In Handbook of Clinical Neurology, M.J. Aminoff, F. Boller, and D.F. Swaab, eds. (Elsevier), pp. 335–356.

Fitzpatrick, D.C., Campbell, A.T., Choudhury, B., Dillon, M.P., Forgues, M., Buchman, C.A., and Adunka, O.F. (2014). Round Window Electrocochleography Just Before Cochlear Implantation: Relationship to Word Recognition Outcomes in Adults. Otol. Neurotol. *35*, 64–71.

Fixmer, E., and Hawkins, S. (1998). The influence of quality of information on the McGurk effect. In AVSP'98 International Conference on Auditory-Visual Speech Processing, p.

Freeman, E.D., Ipser, A., Palmbaha, A., Paunoiu, D., Brown, P., Lambert, C., Leff, A., and Driver, J. (2013). Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing. Cortex *49*, 2875–2887.

Friedland, D.R., Venick, H.S., and Niparko, J.K. (2003). Choice of ear for cochlear implantation: the effect of history and residual hearing on predicted postoperative performance. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *24*, 582–589.

Friesen, L.M., Shannon, R.V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am. *110*, 1150–1163.

Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. (2004). Recalibration of audiovisual simultaneity. Nat. Neurosci. *7*, 773.

Gifford, R.H., Shallop, J.K., and Peterson, A.M. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. Audiol. Neurotol. *13*, 193–205.

Gifford, R.H., Dorman, M.F., Sheffield, S.W., Teece, K., and Olund, A.P. (2014). Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear. Audiol. Neurotol. *19*, 57–71.

Gifford, R.H., Davis, T.J., Sunderhaus, L.W., Menapace, C., Buck, B., Crosson, J., O'Neill, L., Beiter, A., and Segel, P. (2017). Combined Electric and Acoustic Stimulation With Hearing Preservation: Effect of Cochlear Implant Low-Frequency Cutoff on Speech Understanding and Perceived Listening Difficulty. Ear Hear.

Gilley, P.M., Sharma, A., Dorman, M., Finley, C.C., Panch, A.S., and Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. Clin. Neurophysiol. *117*, 1772–1782.

Gilley, P.M., Sharma, A., and Dorman, M.F. (2008). Cortical reorganization in children with cochlear implants. Brain Res. *1239*, 56–65.

Gilley, P.M., Sharma, A., Mitchell, T.V., and Dorman, M.F. (2010). The influence of a sensitive period for auditory-visual integration in children with cochlear implants. Restor. Neurol. Neurosci. *28*, 207–218.

Giraud, A.-L., and Lee, H.-J. (2007). Predicting cochlear implant outcome from brain organisation in the deaf. Restor. Neurol. Neurosci. *25*, 381–390.

Giraud, A.L., Price, C.J., Graham, J.M., Truy, E., and Frackowiak, R.S. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron *30*, 657–663.

Goodwin, J.R., Gaudet, C.R., and Berger, A.J. (2014). Short-channel functional near-infrared spectroscopy regressions improve when source-detector separation is reduced. Neurophotonics *1*, 015002.

Gori, M., Chilosi, A., Forli, F., and Burr, D. (2017). Audio-visual temporal perception in children with restored hearing. Neuropsychologia *99*, 350–359.

Grant, K.W., and Seitz, P.F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. J. Acoust. Soc. Am. *104*, 2438–2450.

Grant, K.W., and Seitz, P.F. (2000a). The use of visible speech cues for improving auditory detection of spoken sentences. J. Acoust. Soc. Am. *108*, 1197–1208.

Grant, K.W., and Seitz, P.-F. (2000b). The use of visible speech cues for improving auditory detection of spoken sentences. J. Acoust. Soc. Am. *108*, 1197–1208.

Grant, K.W., Walden, B.E., and Seitz, P.F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. J. Acoust. Soc. Am. *103*, 2677–2690.

Hassanpour, M.S., Eggebrecht, A.T., Culver, J.P., and Peelle, J.E. (2015). Mapping cortical responses to speech using high-density diffuse optical tomography. NeuroImage *117*, 319–326.

Hay-McCutcheon, M.J., Pisoni, D.B., and Hunt, K.K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. Int. J. Audiol. *48*, 321–333.

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

Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., Ferrill, M., Mickelsen, S., Anderson, S.W., Love, T., et al. (2018). Neural networks supporting audiovisual integration for speech: A large-scale lesion study. Cortex *103*, 360–371.

Hillock, A.R., Powers, A.R., and Wallace, M.T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. Neuropsychologia *49*, 461–467.

Hillock-Dunn, A., and Wallace, M.T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. Dev. Sci. *15*, 688–696.

Hirst, R.J., Stacey, J.E., Cragg, L., Stacey, P.C., and Allen, H.A. (2018). The threshold for the McGurk effect in audio-visual noise decreases with development. Sci. Rep. *8*, 1–12.

Holden, L.K., Finley, C.C., Firszt, J.B., Holden, T.A., Brenner, C., Potts, L.G., Gotter, B.D., Vanderhoof, S.S., Mispagel, K., Heydebrand, G., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. Ear Hear. *34*, 342–360.

Huppert, T.J., Diamond, S.G., Franceschini, M.A., and Boas, D.A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl. Opt. *48*, D280–D298.

Huyse, A., Berthommier, F., and Leybaert, J. (2013). Degradation of labial information modifies audiovisual speech perception in cochlear-implanted children. Ear Hear. *34*, 110–121.

Ipser, A., Agolli, V., Bajraktari, A., Al-Alawi, F., Djaafara, N., and Freeman, E.D. (2017). Sight and sound persistently out of synch: stable individual differences in audiovisual synchronisation revealed by implicit measures of lip-voice integration. Sci. Rep. *7*, 46413.

Irwin, R.J., Ball, A.K.R., Kay, N., Stillman, J.A., and Rosser, J. (1985). The development of auditory temporal acuity in children. Child Dev. *56*, 614.

Jahn, K.N., Stevenson, R.A., and Wallace, M.T. (2017). Visual temporal acuity is related to auditory speech perception abilities in cochlear implant users. Ear Hear. *38*, 236–243.

Kaiser, A.R., Kirk, K.I., Lachs, L., and Pisoni, D.B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. J. Speech Lang. Hear. Res. *46*.

Karns, C.M., Dow, M.W., and Neville, H.J. (2012). Altered Cross-Modal Processing in the Primary Auditory Cortex of Congenitally Deaf Adults: A Visual-Somatosensory fMRI Study with a Double-Flash Illusion. J. Neurosci. *32*, 9626–9638.

Kerlin, J.R., and Shapiro, K.L. (2015). Multisensory Integration: How Sound Alters Sight. Curr. Biol. *25*, R76–R77.

Killion, M.C., Niquette, P.A., Revit, L.J., and Skinner, M.W. (2001). Quick SIN and BKB SIN, two new speech in noise tests permitting SNR 50 estimates in 1 to 2 min. Quick SIN BKB SIN Two New Speech Noise Tests Permitting SNR 50 Estim. 1 2 Min *109*, 2502–2502.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in Psychtoolbox-3. Perception *36*, 1.

Koka, K., Riggs, W.J., Dwyer, R., Holder, J.T., Noble, J.H., Dawant, B.M., Ortmann, A., Valenzuela, C.V., Mattingly, J.K., Harris, M.M., et al. (2018). Intra-Cochlear Electrocochleography During Cochear Implant Electrode Insertion Is Predictive of Final Scalar Location: Otol. Neurotol. *39*, e654–e659.

Kral, A., and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. Trends Neurosci. *35*, 111–122.

Krekelberg, B., and Lappe, M. (2001). Neuronal latencies and the position of moving objects. Trends Neurosci. *24*, 335–339.

Labadie, R.F., Noble, J.H., Hedley-Williams, A.J., Sunderhaus, L.W., Dawant, B.M., and Gifford, R.H. (2016). Results of Postoperative, CT-based, Electrode Deactivation on Hearing in Prelingually Deafened Adult Cochlear Implant Recipients. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *37*, 137–145.

Lachs, L., Pisoni, D.B., and Kirk, K.I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. Ear Hear. *22*, 236–251.

Land, R., Baumhoff, P., Tillein, J., Lomber, S.G., Hubka, P., and Kral, A. (2016). Cross-Modal Plasticity in Higher-Order Auditory Cortex of Congenitally Deaf Cats Does Not Limit Auditory Responsiveness to Cochlear Implants. J. Neurosci. *36*, 6175–6185.

Lazard, D.S., Lee, H.J., Gaebler, M., Kell, C.A., Truy, E., and Giraud, A.L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. NeuroImage *49*, 3443–3451.

Lazard, D.S., Giraud, A.-L., Gnansia, D., Meyer, B., and Sterkers, O. (2012a). Understanding the deafened brain: Implications for cochlear implant rehabilitation. Eur. Ann. Otorhinolaryngol. Head Neck Dis. *129*, 98–103.

Lazard, D.S., Vincent, C., Venail, F., Heyning, P.V. de, Truy, E., Sterkers, O., Skarzynski, P.H., Skarzynski, H., Schauwers, K., O'Leary, S., et al. (2012b). Pre-, Per- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time. PLOS ONE *7*, e48739.

Lazard, D.S., Lee, H.-J., Truy, E., and Giraud, A.-L. (2013). Bilateral reorganization of posterior temporal cortices in post-lingual deafness and its relation to cochlear implant outcome. Hum. Brain Mapp. *34*, 1208–1219.

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

Lee, H., Truy, E., Mamou, G., Sappey-Marinier, D., and Giraud, A. (2007). Visual speech circuits in profound acquired deafness: a possible role for latent multimodal connectivity. Brain *130*, 2929–2941.

Lewis, J.W., and Essen, D.C.V. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J. Comp. Neurol. *428*, 112–137.

Litvak, L.M., Spahr, A.J., Saoji, A.A., and Fridman, G.Y. (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. J. Acoust. Soc. Am. *122*, 982–991.

Lore, W.H., and Song, S. (1991). Central and peripheral visual processing in hearing and nonhearing individuals. Bull. Psychon. Soc. *29*, 437–440.

Love, S.A., Petrini, K., Cheng, A., and Pollick, F.E. (2013). A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments. PLoS ONE *8*, e54798.

Lovelace, C.T., Stein, B.E., and Wallace, M.T. (2003a). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Cogn. Brain Res. *17*, 447–453.

Lovelace, C.T., Stein, B.E., and Wallace, M.T. (2003b). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Cogn. Brain Res. *17*, 447–453.

Macmillan, N.A., and Creelman, C.D. (2004). Detection Theory: A User's Guide (Psychology Press).

Magnotti, J.F., and Beauchamp, M.S. (2015). The noisy encoding of disparity model of the McGurk effect. Psychon. Bull. Rev. *22*, 701–709.

Majdani, O., Leinung, M., Rau, T., Akbarian, A., Zimmerling, M., Lenarz, M., Lenarz, T., and Labadie, R. (2008). Demagnetization of Cochlear Implants and Temperature Changes in 3.0T MRI Environment. Otolaryngol. -- Head Neck Surg. *139*, 833–839.

Mallick, D.B., Magnotti, J.F., and Beauchamp, M.S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. Psychon. Bull. Rev. *22*, 1299–1307.

Marschark, M., and Hauser, P.C. (2008). Deaf cognition: Foundations and outcomes (Oxford University Press).

Marschark, M., Rhoten, C., and Fabich, M. (2007). Effects of Cochlear Implants on Children's Reading and Academic Achievement. J. Deaf Stud. Deaf Educ. *12*, 269–282.

Massaro, D.W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In Categorical Perception: The Groundwork of Cognition, (New York, NY, US: Cambridge University Press), pp. 254–283.

Massaro, D.W., and Cohen, M.M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. J. Speech Lang. Hear. Res. *42*, 21–41.

Massaro, D.W., and Cohen, M.M. (2000). Tests of auditory–visual integration efficiency within the framework of the fuzzy logical model of perception. J. Acoust. Soc. Am. *108*, 784–789.

Massaro, D.W., Thompson, L.A., and Barron, B. (1986). Developmental changes in visual and auditory contributions to speech perception. J. Exp. Child Psychol. 93.

Massaro, D.W., Cohen, M.M., and Thompson, L.A. (1988). Visible Language in Speech Perception: Lipreading and Reading. Visible Lang. Detroit Mich Etc *22*, 8–31.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature *264*, 746–748.

Merabet, L.B., and Pascual-Leone, A. (2010). Neural reorganization following sensory loss: the opportunity of change. Nat. Rev. Neurosci. *11*, 44–52.

Meredith, M.A., and Lomber, S.G. (2011). Somatosensory and visual crossmodal plasticity in the anterior auditory field of early-deaf cats. Hear. Res. *280*, 38–47.

Meredith, M.A., Clemo, H.R., Corley, S.B., Chabot, N., and Lomber, S.G. (2016). Cortical and thalamic connectivity of the auditory anterior ectosylvian cortex of early-deaf cats: Implications for neural mechanisms of crossmodal plasticity. Hear. Res. *333*, 25–36.

Miyamoto, R.T., Osberger, M.J., Todd, S.L., Robbins, A.M., Stroer, B.S., Zimmerman-Phillips, S., and Carney, A.E. (1994). Variables affecting implant performance in children. The Laryngoscope *104*, 1120–1124.

Moody-Antonio, S., Takayanagi, S., Masuda, A., Auer Jr, E.T., Fisher, L., and Bernstein, L.E. (2005). Improved speech perception in adult congenitally deafened cochlear implant recipients. Otol. Neurotol. *26*, 649–654.

Moutoussis, K., and Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. Proc. R. Soc. Lond. B Biol. Sci. *264*, 393–399.

Nadol, J.B., Burgess, B.J., Gantz, B.J., Coker, N.J., Ketten, D.R., Kos, I., Roland, J.T., Shiao, J.Y., Eddington, D.K., Montandon, P., et al. (2001). Histopathology of Cochlear Implants in Humans. Ann. Otol. Rhinol. Laryngol. *110*, 883–891.

Nath, A.R., and Beauchamp, M.S. (2011). Dynamic Changes in Superior Temporal Sulcus Connectivity during Perception of Noisy Audiovisual Speech. J. Neurosci. *31*, 1704–1714.

Nath, A.R., and Beauchamp, M.S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage *59*, 781–787.

Nava, E., Bottari, D., Zampini, M., and Pavani, F. (2008). Visual temporal order judgment in profoundly deaf individuals. Exp. Brain Res. *190*, 179–188.

Neville, H.J., and Lawson, D. (1987a). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. I. Normal hearing adults. Brain Res. *405*, 253–267.

Neville, H.J., and Lawson, D. (1987b). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. II. Congenitally deaf adults. Brain Res. *405*, 268–283.

Niparko, J.K., Tobey, E.A., Thal, D.J., Eisenberg, L.S., Wang, N.-Y., Quittner, A.L., Fink, N.E., and Team, Cd.I. (2010). Spoken language development in children following cochlear implantation. Jama *303*, 1498–1506.

Noble, J.H., Labadie, R.F., Gifford, R.H., and Dawant, B.M. (2013). Image-Guidance Enables New Methods for Customizing Cochlear Implant Stimulation Strategies. IEEE Trans. Neural Syst. Rehabil. Eng. *21*, 820–829.

Noble, J.H., Gifford, R.H., Hedley-Williams, A.J., Dawant, B.M., and Labadie, R.F. (2014). Clinical Evaluation of an Image-Guided Cochlear Implant Programming Strategy. Audiol. Neurotol. *19*, 400–411.

Noble, J.H., Hedley-Williams, A.J., Sunderhaus, L., Dawant, B.M., Labadie, R.F., Camarata, S.M., and Gifford, R.H. (2016). Initial Results With Image-guided Cochlear Implant Programming in Children. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *37*, e63-69.

Olds, C., Pollonini, L., Abaya, H., Larky, J., Loy, M., Bortfeld, H., Beauchamp, M.S., and Oghalai, J.S. (2015). Cortical Activation Patterns Correlate With Speech Understanding After Cochlear Implantation. Ear Hear.

Olds, C., Pollonini, L., Abaya, H., Larky, J., Loy, M., Bortfeld, H., Beauchamp, M.S., and Oghalai, J.S. (2016). Cortical activation patterns correlate with speech understanding after cochlear implantation. Ear Hear. *37*, e160–e172.

Pantev, C., Dinnesen, A., Ross, B., Wollbrink, A., and Knief, A. (2006). Dynamics of Auditory Plasticity after Cochlear Implantation: A Longitudinal Study. Cereb. Cortex *16*, 31–36.

Parasnis, I., and Samar, V.J. (1985). Parafoveal attention in congenitally deaf and hearing young adults. Brain Cogn. *4*, 313–327.

Pascual-Leone, A., and Hamilton, R. (2001). Chapter 27 The metamodal organization of the brain. Prog. Brain Res. *134*, 427–445.

Peters, B.R., Wyss, J., and Manrique, M. (2010). Worldwide trends in bilateral cochlear implantation. The Laryngoscope *120*, S17–S44.

Peterson, G.E., and Lehiste, I. (1962). Revised CNC lists for auditory tests. J. Speech Hear. Disord. *27*, 62–70.

Petitto, L.A., Zatorre, R.J., Gauna, K., Nikelski, E.J., Dostie, D., and Evans, A.C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. Proc. Natl. Acad. Sci. *97*, 13961–13966.

Picou, E.M., and Ricketts, T.A. (2014). The Effect of Changing the Secondary Task in Dual-Task Paradigms for Measuring Listening Effort: Ear Hear. *35*, 611–622.

Picou, E.M., Ricketts, T.A., and Hornsby, B.W.Y. (2011). Visual Cues and Listening Effort: Individual Variability. J. Speech Lang. Hear. Res. *54*, 1416–1430.

Picou, E.M., Charles, L.M., and Ricketts, T.A. (2017). Child–Adult Differences in Using Dual-Task Paradigms to Measure Listening Effort. Am. J. Audiol. *26*, 143–154.

Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M.S., and Oghalai, J.S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. Hear. Res. *309*, 84–93.

Powers, A.R., Hillock, A.R., and Wallace, M.T. (2009). Perceptual training narrows the temporal window of multisensory binding. J. Neurosci. *29*, 12265–12274.

Powers III, A.R., Hillock-Dunn, A., and Wallace, M.T. (2016). Generalization of multisensory perceptual learning. Sci. Rep. *6*.

Puschmann, S., and Thiel, C.M. (2017). Changed crossmodal functional connectivity in older adults with hearing loss. Cortex *86*, 109–122.

Quaresima, V., Bisconti, S., and Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. Brain Lang. *121*, 79–89.

Quinto, L., Forde Thompson, W., Russo, F.A., and Trehub, S.E. (2010). A comparison of the McGurk effect for spoken and sung syllables. Atten. Percept. Psychophys. *72*, 1450–1454.

Reich, L., Szwed, M., Cohen, L., and Amedi, A. (2011). A Ventral Visual Stream Reading Center Independent of Visual Experience. Curr. Biol. *21*, 363–368.

Reynolds, H.N. (1993). Effects of Foveal Stimulation on Peripheral Visual Processing and Laterality in Deaf and Hearing Subjects. Am. J. Psychol. *106*, 523–540.

Riggs, W.J., Dwyer, R.T., Holder, J.T., Mattingly, J.K., Ortmann, A., Noble, J.H., Dawant, B.M., Valenzuela, C.V., O'Connell, B.P., Harris, M.S., et al. (2019). Intracochlear Electrocochleography: Influence of Scalar Position of the Cochlear Implant Electrode on Postinsertion Results. Otol. Neurotol. *40*, e503–e510.

van de Rijt, L.P.H., van Opstal, A.J., Mylanus, E.A.M., Straatman, L.V., Hu, H.Y., Snik, A.F.M., and van Wanrooij, M.M. (2016). Temporal Cortex Activation to Audiovisual Speech in Normal-Hearing and Cochlear Implant Users Measured with Functional Near-Infrared Spectroscopy. Front. Hum. Neurosci. *10*.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., and Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb. Cortex N. Y. N 1991 *17*, 1147–1153.

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

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

Rubinstein, J.T., Parkinson, W.S., Tyler, R.S., and Gantz, B.J. (1999). Residual speech recognition and cochlear implant performance: effects of implantation criteria. Am. J. Otol. *20*, 445–452.

Runge, C.L., Henion, K., Tarima, S., Beiter, A., and Zwolan, T.A. (2016). Clinical Outcomes of the Cochlear™ Nucleus® 5 Cochlear Implant System and SmartSound™ 2 Signal Processing. J. Am. Acad. Audiol. *27*, 425–440.

Sakatani, K., Xie, Y., Lichty, W., Li, S., and Zuo, H. (1998). Language-activated cerebral blood oxygenation and hemodynamic changes of the left prefrontal cortex in poststroke aphasic patients: a near-infrared spectroscopy study. Stroke *29*, 1299–1304.

Saliba, J., Bortfeld, H., Levitin, D.J., and Oghalai, J.S. (2016). Functional near-infrared spectroscopy for neuroimaging in cochlear implant recipients. Hear. Res.

Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R.D., Marcar, V.L., Jäncke, L., and Debener, S. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. Brain 555–568.

Schneider, K.A., and Bavelier, D. (2003). Components of visual prior entry. Cognit. Psychol. *47*, 333–366.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-Prime: User's guide (Psychology Software Incorporated).

Scholkmann, F., Kleiser, S., Metz, A.J., Zimmermann, R., Mata Pavia, J., Wolf, U., and Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. NeuroImage *85*, 6–27.

Schorr, E.A., Fox, N.A., van Wassenhove, V., and Knudsen, E.I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. Proc. Natl. Acad. Sci. U. S. A. *102*, 18748–18750.

Schreitmüller, S., Frenken, M., Bentz, L., Ortmann, M., Walger, M., and Meister, H. (2017). Validating a method to assess lipreading, audiovisual gain, and integration during speech reception with cochlear-implanted and normal-hearing subjects using a talking head. Ear Hear.

Schwartz, J.-L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. J. Acoust. Soc. Am. *127*, 1584–1594.

Seltzer, B., Cola, M.G., Gutierrez, C., Massee, M., Weldon, C., and Cusick, C.G. (1996). Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. J. Comp. Neurol. *370*, 173–190.

Sevy, A.B.G., Bortfeld, H., Huppert, T.J., Beauchamp, M.S., Tonini, R.E., and Oghalai, J.S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. Hear. Res. *270*, 39–47.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. Nature *408*, 788.

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. Cogn. Brain Res. *14*, 147–152.

Shannon, R.V. (1989). Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. J. Acoust. Soc. Am. *85*, 2587–2592.

Sharma, A., Nash, A.A., and Dorman, M. (2009). Cortical development, plasticity and re-organization in children with cochlear implants. J. Commun. Disord. *42*, 272–279.

Shiell, M.M., Champoux, F., and Zatorre, R.J. (2016). The Right Hemisphere Planum Temporale Supports Enhanced Visual Motion Detection Ability in Deaf People: Evidence from Cortical Thickness. Neural Plast. *2016*, 1–9.

Skottun, B.C., and Skoyles, J.R. (2010). Temporal order judgment in dyslexia—Task difficulty or temporal processing deficiency? Neuropsychologia *48*, 2226–2229.

Sommers, M.S., Tye-Murray, N., and Spehar, B. (2005). Auditory-Visual Speech Perception and Auditory-Visual Enhancement in Normal-Hearing Younger and Older Adults: Ear Hear. *26*, 263–275.

Spahr, A.J., Dorman, M.F., Litvak, L.M., Van Wie, S., Gifford, R.H., Loizou, P.C., Loiselle, L.M., Oakes, T., and Cook, S. (2012). Development and validation of the AzBio sentence lists. Ear Hear. *33*, 112–117.

Spence, C., and Parise, C. (2010). Prior-entry: A review. Conscious. Cogn. *19*, 364–379.

Stein, B.E., London, N., Wilkinson, L.K., and Price, D.D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. J. Cogn. Neurosci. *8*, 497–506.

Stevenson, R.A., and James, T.W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. NeuroImage *44*, 1210–1223.

Stevenson, R.A., Zemtsov, R.K., and Wallace, M.T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. J. Exp. Psychol. Hum. Percept. Perform. *38*, 1517–1529.

Stevenson, R.A., Siemann, J.K., Woynaroski, T.G., Schneider, B.C., Eberly, H.E., Camarata, S.M., and Wallace, M.T. (2014a). Evidence for Diminished Multisensory Integration in Autism Spectrum Disorders. J. Autism Dev. Disord. *44*, 3161–3167.

Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.R., Siemann, J.K., James, T.W., and Wallace, M.T. (2014b). Identifying and Quantifying Multisensory Integration: A Tutorial Review. Brain Topogr. *27*, 707–730.

Stevenson, R.A., Siemann, J.K., Schneider, B.C., Eberly, H.E., Woynaroski, T.G., Camarata, S.M., and Wallace, M.T. (2014c). Multisensory Temporal Integration in Autism Spectrum Disorders. J. Neurosci. *34*, 691–697.

Stevenson, R.A., Sheffield, S.W., Butera, I.M., Gifford, R.H., and Wallace, M.T. (2017a). Multisensory integration in cochlear implant recipients. Ear Hear. *38*, 521–538.

Stevenson, R.A., Baum, S.H., Krueger, J., Newhouse, P.A., and Wallace, M.T. (2017b). Links Between Temporal Acuity and Multisensory Integration Across Life Span. J. Exp. Psychol. Hum. Percept. Perform.

Stevenson, R.A., Park, S., Cochran, C., McIntosh, L.G., Noel, J.-P., Barense, M.D., Ferber, S., and Wallace, M.T. (2017c). The associations between multisensory temporal processing and symptoms of schizophrenia. Schizophr. Res. *179*, 97–103.

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

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Démonet, J.-F., Déguine, O., and Barone, P. (2015b). Increased audiovisual integration in cochlear-implanted deaf patients: independent components analysis of longitudinal positron emission tomography data. Eur. J. Neurosci. *41*, 677–685.

Stropahl, M., and Debener, S. (2017). Auditory cross-modal reorganization in cochlear implant users indicates audio-visual integration. NeuroImage Clin. *16*, 514–523.

Stropahl, M., Plotz, K., Schönfeld, R., Lenarz, T., Sandmann, P., Yovel, G., De Vos, M., and Debener, S. (2015). Cross-modal reorganization in cochlear implant users: Auditory cortex contributes to visual face processing. NeuroImage *121*, 159–170.

Stropahl, M., Chen, L.-C., and Debener, S. (2017a). Cortical reorganization in postlingually deaf cochlear implant users: Intra-modal and cross-modal considerations. Hear. Res. *343*, 128–137.

Stropahl, M., Schellhardt, S., and Debener, S. (2017b). McGurk stimuli for the investigation of multisensory integration in cochlear implant users: The Oldenburg Audio Visual Speech Stimuli (OLAVS). Psychon. Bull. Rev. *24*, 863–872.

Sumby, W.H., and Pollack, I.P. (1954). Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. *26*, 212–215.

Tan, L., Holland, S.K., Deshpande, A.K., Chen, Y., Choo, D.I., and Lu, L.J. (2015). A semi-supervised Support Vector Machine model for predicting the language outcomes following cochlear implantation based on pre-implant brain fMRI imaging. Brain Behav. *5*, e00391.

Tiippana, K., Andersen, T.S., and Sams, M. (2004). Visual attention modulates audiovisual speech perception. Eur. J. Cogn. Psychol. *16*, 457–472.

Tona, R., Naito, Y., Moroto, S., Yamamoto, R., Fujiwara, K., Yamazaki, H., Shinohara, S., and Kikuchi, M. (2015). Audio–visual integration during speech perception in prelingually deafened Japanese children revealed by the McGurk effect. Int. J. Pediatr. Otorhinolaryngol. *79*, 2072–2078.

Tremblay, C., Champoux, F., Voss, P., Bacon, B.A., Lepore, F., and Théoret, H. (2007). Speech and Non-Speech Audio-Visual Illusions: A Developmental Study. PLoS ONE *2*, e742.

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

Tye-Murray, N., Sommers, M.S., and Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear Hear. *28*, 656–668.

Tyler, R.S., Fryauf-Bertschy, H., Kelsay, D.M., Gantz, B.J., Woodworth, G.P., and Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. Otolaryngol. Head Neck Surg. *117*, 180–187.

Uludağ, K., Steinbrink, J., Villringer, A., and Obrig, H. (2004). Separability and cross talk: optimizing dual wavelength combinations for near-infrared spectroscopy of the adult head. NeuroImage *22*, 583–589.

Van Engen, K.J., Xie, Z., and Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. Atten. Percept. Psychophys. *79*, 396–403.

Van Wassenhove, V., Grant, K.W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. U. S. A. *102*, 1181–1186.

Vatakis, A., and Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. Neurosci. Lett. *393*, 40–44.

Violentyev, A., Shimojo, S., and Shams, L. (2005). Touch-induced visual illusion. Neuroreport *16*, 1107–1110.

Voss, P. (2013). Sensitive and critical periods in visual sensory deprivation. Front. Psychol. *4*.

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

Vroomen, J., and Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. Atten. Percept. Psychophys. *72*, 871–884.

Wagner, L., Maurits, N., Maat, B., Başkent, D., and Wagner, A.E. (2018). The Cochlear Implant EEG Artifact Recorded From an Artificial Brain for Complex Acoustic Stimuli. IEEE Trans. Neural Syst. Rehabil. Eng. *26*, 392–399.

Wallace, M.T., and Stevenson, R.A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. Neuropsychologia *64*, 105–123.

Wang, N.-Y., Eisenberg, L.S., Johnson, K.C., Fink, N.E., Tobey, E.A., Quittner, A.L., and Niparko, J.K. (2008). Tracking Development of Speech Recognition: Longitudinal Data From Hierarchical Assessments in the Childhood Development After Cochlear Implantation Study. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *29*, 240–245.

Wanna, G.B., Noble, J.H., Carlson, M.L., Gifford, R.H., Dietrich, M.S., Haynes, D.S., Dawant, B.M., and Labadie, R.F. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes: Impact of Electrode Type and Approach on CI Outcome. The Laryngoscope *124*, S1–S7.

van Wassenhove, V., Grant, K.W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia *45*, 598–607.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. NeuroImage *31*, 1247–1256.

Wiggins, I.M., and Hartley, D.E.H. (2015). A Synchrony-Dependent Influence of Sounds on Activity in Visual Cortex Measured Using Functional Near-Infrared Spectroscopy (fNIRS). PLOS ONE *10*, e0122862.

Wilson, B., and Dorman, M.F. (2012). Better hearing with cochlear implants: Studies at the research triangle institute (Plural Publishing).

Woynaroski, T.G., Kwakye, L.D., Foss-Feig, J.H., Stevenson, R.A., Stone, W.L., and Wallace, M.T. (2013). Multisensory Speech Perception in Children with Autism Spectrum Disorders. J. Autism Dev. Disord. *43*, 2891–2902.

Yamamoto, R., Naito, Y., Tona, R., Moroto, S., Tamaya, R., Fujiwara, K., Shinohara, S., Takebayashi, S., Kikuchi, M., and Michida, T. (2017). Audio-visual speech perception in prelingually deafened Japanese children following sequential bilateral cochlear implantation. Int. J. Pediatr. Otorhinolaryngol. *102*, 160–168.

Zampini, M., Shore, D.I., and Spence, C. (2005). Audiovisual prior entry. Neurosci. Lett. *381*, 217–222.

Zuniga, M.G., Rivas, A., Hedley-Williams, A., Gifford, R.H., Dwyer, R., Dawant, B.M., Sunderhaus, L.W., Hovis, K.L., Wanna, G.B., Noble, J.H., et al. (2017). Tip fold-over in cochlear implantation: case series. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *38*, 199–206.