8-13-2021 2:00 PM

Audiovisual integration in cochlear implant users and typical hearing controls: A study of group differences in syllable perception and effect of asynchrony on speech intelligibility

Cailey A. Salagovic, *The University of Western Ontario*

Supervisor: Butler, Blake E., *The University of Western Ontario*
Co-Supervisor: Stevenson, Ryan A., *University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Psychology

Abstract

This study examined audiovisual integration in cochlear implant (CI) users compared to typical (acoustic) hearing control participants and investigated the effect of audiovisual temporal asynchrony on speech intelligibility across these groups. Additionally, this study evaluated the utility of online data collection for audiovisual perception research. In Experiment 1, CI users were found to integrate audiovisual syllables comparably to controls as demonstrated by perception of the McGurk illusion. However, group differences were revealed in the processing of the unisensory components and underlying distributions of responses to incongruent audiovisual trials when the illusory fusion syllable was not reported. In Experiment 2, intelligibility of sentences presented in noise was more facilitated by the presence of visual cues and more inhibited by temporal offset for CI users than controls. Together these results indicate a functionally relevant difference in how CI users process and combine auditory and visual speech signals compared to control participants.

Keywords

Multisensory Perception, Audiovisual Integration, Cochlear Implants, McGurk, Speech-in-Noise, Audiovisual Asynchrony, Online Data Collection

Summary for Lay Audience

When a person is seen speaking, our ability to understand their speech is supported by both the sound of the voice and visual cues arising from mouth movements. The relative amount that these auditory and visual cues contribute to understanding these multisensory signals varies depending on the situation. For instance, in noisy environments listeners watch a talker's mouth closely to compensate for difficulty hearing their voice. People who use cochlear implants (CIs), hearing devices that bypass damaged regions of the ear to convey auditory information directly to the brain, may have a similar experience. Because the auditory signal produced by CIs is less clear than that conveyed by the typically-developed inner ear, CI users rely on visual speech cues more than those with typical hearing. The goal of this study was to investigate audiovisual integration in CI users compared to typical hearing controls and evaluate how audiovisual asynchrony affects speech comprehension in these groups. Experiment 1 used the McGurk illusion in which a speaker's mouth is seen to say one syllable, like "ba", while their voice is heard to say a different syllable, like "ga". Because the brain automatically integrates audiovisual speech information, many people experience an illusory syllable, like "da", that represents a fusion of the auditory and visual information. We found that CI users experience this illusion at a rate comparable to control participants. However, when they didn't experience the illusion, CI users usually reported the seen syllable whereas control participants reported the heard syllable. In Experiment 2, participants watched videos of sentences spoken in background noise and typed what they heard. The sound and video were aligned for some sentences, and out of synch for others. The addition of visual cues enhanced accuracy more for CI users than control participants. CI users' accuracy was also more inhibited by asynchrony than control participants. These findings indicate that CI users combine auditory and visual speech information differently than individuals with typical hearing and these differences affect CI users' ability to understand asynchronous speech. This is pertinent given the increasing use of teleconferencing platforms, which are prone to audiovisual lag.

Acknowledgments

I would like to thank my supervisors Dr. Blake Butler and Dr. Ryan Stevenson for all their guidance and support throughout the course of this project. I would also like to thank Dr. Rebecca Hirst for her invaluable training and assistance in bringing this project to Pavlovia. I am grateful to the many participants who graciously contributed their time and effort, without whom this work would not have been possible. Finally, my most sincere gratitude is owed to my parents and my friends for always believing in me.

Table of Contents

List of Tables

List of Abbreviations

CI user	Individual with cochlear implant(s)
d	Cohen’s d effect size value
dB	Decibel
F	F distribution value
M	Mean
ms	Millisecond(s)
n	Number of participants in group
p	Probability value
SD	Standard Deviation
SNR	Signal-to-noise ratio
U	Mann-Whitney U statistic value

List of Figures

List of Appendices

Chapter 1

1 Introduction

1.1 Audiovisual Speech Perception

Given the semantically, socially, and acoustically rich content conveyed by the human voice, spoken language is often considered to be a primarily auditory construct. Indeed, auditory-only speech is often fully intelligible without any visual cues. Yet, when the person speaking can be seen, the mouth and face movements involved in speech production provide a nuanced visual element to spoken language (Rosenblum & Saldaña, 1996). In environments both with and without background noise, this visual information plays an important role in speech perception. Sumby & Pollack (1954) famously demonstrated that, when competing background noise is present, listeners are better able to understand what is being said when they can see the talker. These compelling findings are further supported by the near universal experience of following spoken language, whether with ease or some degree of difficulty, across a variety of real-world environments, from private conversations in a quiet home to snatches of social banter at a noisy cocktail party.

Thus, spoken language is an inherently audiovisual phenomenon. When available, complementary auditory and visual streams are integrated into a perceptually unified, multisensory signal. The brain is adapted to efficiently combine stimuli from multiple sensory modalities to best detect and respond to objects and events of relevance in the environment. Multisensory stimuli are especially salient, as the integrated percept can provide more information about the precipitating external event more quickly than either of the unisensory components alone or in sum (Stein & Stanford, 2008). While many environmental events can be perceived as multisensory phenomena, speech signals appear to be processed uniquely, such that these stimuli are especially likely to be integrated (Tuomainen *et al.*, 2005).

The McGurk illusion (McGurk & MacDonald, 1976) is a widely-cited demonstration of the audiovisual nature of speech. In this illusion, a speaker’s voice is heard to utter one syllable while the mouth is seen to produce a different syllable, giving rise to an illusory

percept of a third, ‘fused’ syllable. For instance, the combination of an aurally-presented /ba/ and visually-presented /ga/ is often perceived as the illusory syllable /da/. In this case, /da/ is interpreted as an intermediate syllable that fuses together phonemic elements of both presented syllables into a new percept. However, the definition of a ‘fused’ syllable varies across the McGurk literature with some studies specifying that only certain response syllables that have phonemic characteristics of both presented syllables represent true fusion whereas others consider any response other than the aurally-presented syllable to qualify as perception of the illusion (Getz & Toscano, 2021). McGurk and MacDonald (1976) interpreted their finding as an indication that multisensory transformation occurs such that the input from the two modalities becomes an entirely new percept, no element of which need be present in the original stimuli. That the brain is inclined to derive a singular, illusory percept from disparate unisensory signals suggests that the processing of concurrent auditory and visual speech information and the integration of these signals is automatic in the perceptual processing of real-world sensory signals.

Studies of the McGurk illusion have shown that illusory fused syllables can be evoked from a variety of syllable combinations (McGurk & Macdonald, 1976; Stropahl *et al.*, 2017). The illusion has been produced across a variety of age groups and languages, with various manipulations of the stimuli, and is even robust to knowledge of the illusion, though the strength of the illusion does vary across these parameters and between individuals (Rosenblum, 2019). Factors such as specific stimulus features, task design, and participant characteristics that are unrelated to the actual perceptual information provided by the stimuli introduce substantial variability in the likelihood that an observer will experience the illusion (Getz & Toscano, 2021). While this variability suggests the McGurk illusion may not reflect a fundamental construct of audiovisual integration, the illusion is well-suited for research investigating the relative weighting of the auditory and visual modalities in the processing of speech stimuli (Getz & Toscano, 2021).

Where present, visual speech cues impact perception even when auditory signals alone are sufficient for basic intelligibility. However, perceptual gain related to multisensory processing has been shown to be greatest when the contributing unisensory

inputs are weak (Stein & Stanford, 2008). This *principle of inverse effectiveness* is established at the level of individual multisensory neurons, in which activation is greatest when a weak unisensory signal is enhanced through combination with a signal from another modality. This increase in the rate of action potential generation is greater in response to a minimally salient stimulus (i.e. one just above the detection threshold) than for a highly salient unisensory stimulus. Thus, there is an inverse relationship between the effectiveness of the individual stimulus and the extent of the perceptual enhancement gained from the addition of a second modality. Through this merging of sensory inputs, stimuli that may be ineffective alone can significantly alter the efficacy of other stimuli, serving to strengthen the effect of otherwise faint environmental cues (Meredith & Stein, 1983). This principle is most clearly observed in cells of the superior colliculus, an area which is understood to be specialized for detection of sensory signals and orientation toward their environmental sources. Corresponding behavioural research has shown that reaction times to multisensory stimuli are faster than to unisensory signals (Diederich & Colonius, 2004).

While detection and orientation are certainly involved in speech perception, the relationship between signal strength and integration for speech perception is thought to be driven by higher order semantic processing. In challenging listening environments, integration of audiovisual speech has been shown to occur at a linguistic level, over longer temporal windows (Crosse *et al.*, 2016). Due to the spectrotemporal complexity of the signal, the classic inverse effectiveness relationship may not apply to audiovisual speech stimuli. Ross *et al.* (2007) tested typical hearing listeners with open-ended word identification across seven signal-to-noise ratios (SNRs) ranging from 0 dB to -24 dB and found that, while audio-alone intelligibility is present at SNRs as low as -20 dB, the greatest multisensory facilitation arose in an intermediate range, around -12 dB SNR. Similarly, in behavioural pilot testing, Crosse *et al.* (2016) identified an intermediate SNR of -9 dB as the point of optimal multisensory gain for continuous speech stimuli. This “special zone” of multisensory enhancement is unique; at these SNRs, both unisensory stimuli provide poor speech comprehension, so intelligibility relies on accessing and integrating the available cues. By contrast, at higher SNRs, the system can function with near complete reliance on the auditory stream, while at lower SNRs, any

comprehension achieved is likely derived from the visual stream alone via lip reading (Ross *et al.*, 2007).

Regardless of the exact relationship between signal strength and multisensory integration, it is clear that the visual component of spoken language is meaningful, and especially useful when the complementary auditory signal is degraded in some way. This raises the question of how the relative weighting of these signals is affected in scenarios where all incoming auditory input is fundamentally degraded, as is the case for individuals who use cochlear implants (CI users).

1.2 Audiovisual Speech Perception in Cochlear Implant Users

In 2018, the World Health Organization recognized hearing loss as the fourth highest cause of disability globally. This number is expected to increase significantly in coming years in accordance with demographic trends toward a growing and aging world population. In addition, over one billion young people are at risk of developing hearing loss due to widespread access to smart phones and increasing duration and volume of music listening. If current trends continue, an estimated 630 million people will be living with disabling hearing loss by 2030 (WHO, 2018).

Cochlear implantation is an increasingly common and highly effective procedure used to provide a sense of sound to individuals with sensorineural deafness. Unlike a hearing aid, which amplifies and, in some cases, shifts the frequency content of incoming sound so as to be detected by intact structures of the inner ear, a cochlear implant is a neuroprosthesis that bypasses damage in the ear to stimulate the auditory nerve directly. Externally, a microphone is worn behind the ear which captures environmental sounds. An analog waveform representation of these sounds is then conveyed to an external speech processor which translates that waveform into a pattern of electrical stimulation. A transmitter on the outside of the scalp receives this processed signal and sends it to a receiver implanted under the scalp. This receiver also acts as a stimulator which sends electrical signals to a number of electrode contacts arranged along a thin, flexible array inserted into the cochlea. These electrode contacts are placed in close proximity to the basilar membrane where they transmit the processed stimulation as electrical impulses to spiral ganglion neurons. Finally, from the auditory nerve these signals are transmitted via

the ascending auditory pathway to auditory cortex where they are perceived as sound (Moore & Carlyon, 2005; NIDCD, 2021; Yawn *et al.*, 2015).

Cochlear implants are considered the most successful neuroprosthesis available to modern healthcare and have been shown to significantly improve hearing-specific and overall quality of life for users, including reduction in experiences of isolation, depression, and functional limitations associated with hearing loss (Buchman *et al.*, 2020). As of December 2019, the National Institute on Deafness and Other Communication Disorders reports that more that 736,000 registered devices have been implanted worldwide (NIDCD, 2021). Conventionally, cochlear implants are offered to pre-school aged children through adults following the general criteria that candidates have bilateral profound deafness or severe hearing loss and are not benefited by hearing aids. Candidacy criteria are continuing to expand with options becoming available for children younger than one year of age, individuals with unilateral hearing loss, and those with residual low frequency hearing (Yawn *et al.*, 2015). The implantation itself is a low-risk outpatient procedure and activation of the device usually occurs two to four weeks following implantation. At this activation, an audiologist will work with the user to calibrate each electrode contact to give the individual the greatest possible range of frequency representation and ensure the output of the device is audible but does not reach sound levels that could elicit pain.

The signals provided by cochlear implants provide a useful representation of external sounds that can help users to understand speech and other environmental noises. However, this sensation should not be construed as restoration of acoustic hearing. The inner ear and its innervation into auditory cortex represent a sophisticated system comprised of highly specialized structures and mechanisms that work in tandem to capture, relay, and process complex soundscapes. Though cochlear implants are decidedly effective sensory prostheses, the full function of these devices is limited by how they interface with the existing hearing system. Limitations arise primarily from the physical properties and placement of the electrodes. In acoustic hearing, there is a nearly one-to-one mapping of fine-grained frequency information between inner hair cells and individual spiral ganglion neurons. This specificity cannot be replicated by cochlear implants as each electrode contact typically interfaces with a small population of neurons.

With regards to placement, the arrangement of the electrodes mimics the pitch-mapping within the typically-developed cochlea with high frequencies being transmitted by electrodes near the base of the cochlea progressing through low frequencies conveyed by electrodes nearer to the apex. However, the physical constraints of the electrode array itself often preclude full insertion into the apex of the cochlea, such that auditory nerve fibers that are typically activated by low frequencies present in speech and other common environmental noises are not accessed. Due to this limitation, the low frequency signals passed by the device are relayed by electrodes interfacing with nerve fibers that would normally be tuned to higher frequencies. Thus, a there is a fundamental mismatch between stimulated frequencies and the natural tonotopic tuning of the auditory system. Within the scope of these limitations, efforts are made in initial programming of the device to align the frequency stimulated by each electrode as closely as possible to typical pitch mapping. Programming also aims to avoid cross stimulation of discrete neural populations by multiple electrodes. In many individuals with hearing loss, some areas of the cochlea may be fully degenerated, producing "dead zones" that cannot be stimulated with any frequency. For these reasons, even in successful implantations, frequency-to-place mappings may be misaligned up to 3 octaves, severely affecting the representation of incoming sounds (Moore & Carlyon, 2005). As a result, individuals who receive cochlear implants as adults must actively learn to interpret the sounds generated by the device over time through practice and guided therapies (NIDCD, 2021).

Despite degradation of the auditory signal, a majority of CI users have good clinical outcomes, most commonly defined as success using the device to understand spoken language in quiet. In postlingually deafened individuals who received unilateral cochlear implants as adults, comprehension of syllables, single words, and full sentences in quiet improved significantly after implantation regardless of age at implantation (Lachowska *et al*., 2014; Park *et al.*, 2011). However, understanding speech in the presence of background noise remains a widely reported challenge, regardless of individual proficiency in quiet (Fetterman & Domico, 2002; Hochberg *et al.*, 1992). Depending upon the acoustic characteristics of the background noise and the speech stimuli, speech reception thresholds of CI users have been shown to be 10 – 25 dB higher than those with acoustic hearing (Spriet *et al.*, 2007; Zeng *et al.*, 2005).

Speech perception gains after implantation, whether in quiet or noise, are generally measured with auditory-only stimuli. Although literature examining multisensory integration in CI users is relatively scarce, it is agreed that CI users are able to integrate input from auditory and visual modalities, and do show behavioural benefits associated with multisensory perceptual gain. The extent to which any individual CI user may be able to effectively integrate auditory and visual signals depends on a number of factors including age of implantation, duration of deafness, and the demands of the assessment task being used (Stevenson *et al.*, 2017). Findings regarding the extent of audiovisual integration in CI users compared to those observed in acoustic hearing are mixed. Using a combination of reaction time measures for consonant identification and accuracy measures for consonant-nucleus-consonant words, Zhou *et al.* (2019) found that CI users showed similar, but not better, audiovisual integration than acoustic hearing controls. Conversely, using disyllabic words, Rouger *et al.* (2007), found that CI users were better able to integrate visual and auditory speech information than acoustic hearing controls for whom audiovisual stimuli were created that simulated the signal provided by a cochlear implant.

Given that the auditory input provided by a cochlear implant is fundamentally degraded relative to that provided by the intact cochlea, CI users are likely to rely on visual speech information to a greater extent than those with acoustic hearing (Desai *et al.*, 2008). Thus, questions are raised regarding the relative weightings of the auditory and visual streams in the integration of speech by CI users. Again, the McGurk illusion offers the means to examine the relative roles of these two sensory modalities in the perception of a unified speech percept. Because the perception of an illusory 'fused' syllable is dependent on myriad stimulus and listener features, a single individual rarely perceives fusion on every trial of a McGurk-style experiment. When a non-fused percept is reported, the perceived syllable typically corresponds to that which was heard or seen by the participant; accordingly, on such trials the relative rates at which the auditory or visual component is reported can provide as estimate of which sensory modality is contributing most strongly to speech perception. Studies of the McGurk illusion in CI users have shown that the illusory fused syllables are perceived less frequently than by acoustic hearing controls. Furthermore, when the illusion is not perceived, CI users are

more likely to report the visually-presented syllable whereas individuals with acoustic hearing tend to report the aurally-presented syllable (see Stevenson *et al.*, 2017 for review). More recently, in validating a highly normalized set of McGurk stimuli (the Oldenburg Audio Visual Speech Stimuli [OLAVS]) and applying a probabilistic model accounting for individual and stimulus-dependent parameters, Stropahl *et al.* (2017) reported *higher* rates of overall fusion perception in CI users than in their acoustic hearing control group. In discussing their findings, the authors suggest that stimulus-effects present in other studies of McGurk perception in CI users make it difficult to compare outcomes across these studies and suggest that the OLAVS set allows for better capture of existing group-level differences in illusory perception.

At present, speech perception in CI users is primarily quantified by clinical evaluations that present auditory only speech sounds in highly controlled acoustic environments (i.e. sound-attenuating booths; Sargent *et al.*, 2001). Additionally, research addressing multisensory integration in this population is relatively scarce and often relies on various measures of audiovisual integration which do not correlate well with each other, and thus do not likely reflect the same underlying integrative processes (Wilbiks *et al.*, 2021). Neither these clinical nor research scenarios reflect the noisy, dynamic, multisensory, real-world environments in which cochlear implants are primarily used. Better understanding of the real-world functionality and limitations of these devices is extremely prudent as listening environments including large classrooms, open-plan office spaces, and virtual working and learning environments continue to become more prevalent. These difficult listening scenarios require consideration of background noise, perceptual gain from visual cues, and issues related to degraded audiovisual stimuli inherent to online video calling platforms. Thorough understanding of cochlear implant performance in these environments is crucial for developing an accurate depiction of how individuals are using their devices and could contribute to better outcomes overall through more targeted therapeutic and signal processing options.

1.3 Asynchrony in Audiovisual Speech Perception

The extent to which unisensory stimuli are combined to form an integrated percept is dependent on multiple factors including spatial and temporal alignment (Stein &

Stanford, 2008). The more closely together in time an auditory and visual event occur, the more likely they are to be bound into a multisensory unit. The range of temporal offsets over which perception and subsequent response is most likely to be enhanced by the presence of two unisensory signals is known as the *temporal binding window* (Wallace & Stevenson, 2014). In research, this temporal binding window is often used as a proxy for the occurrence of multisensory integration.

The temporal binding window is typically illustrated as a probability curve (see Figure 1), with a peak at the audiovisual offset where integration is most likely to occur - an individual's *point of subjective simultaneity*. On either side of this peak, the curve gradually declines as integration becomes less probable at greater degrees of offset. The curve is commonly asymmetrical with the peak occurring when the visual signal occurs slightly before the auditory signal. The slope of the decline may also be skewed such that greater levels of integration are maintained through larger offsets in conditions when the visual stream precedes the auditory stream (Vroomen & Keetels, 2010). Ideally, the range of offsets within an individual's temporal binding window is not so wide as to mischaracterize events that are truly sequential as co-occurring, and not so narrow as to preclude the binding of related signals due to natural temporal differences in stimulus detection and processing (e.g. the small-scale differences between arrival time that result from the difference between the speeds of light and sound).

The specific characteristics, including overall width and shape, of the temporal binding window vary naturally across individuals (Stevenson *et al.*, 2012) and are likely to be affected by hearing experience. A period of hearing loss, especially during critical periods of development, followed by a period of experience with implant-generated sounds likely contributes to differentiation in temporal binding window characteristics between CI users and individuals with acoustic hearing. Greater flexibility in the range of asynchronies over which audiovisual integration occurs may be adaptive in CI users, allowing the perceptual gains associated with integration to be experienced more readily.

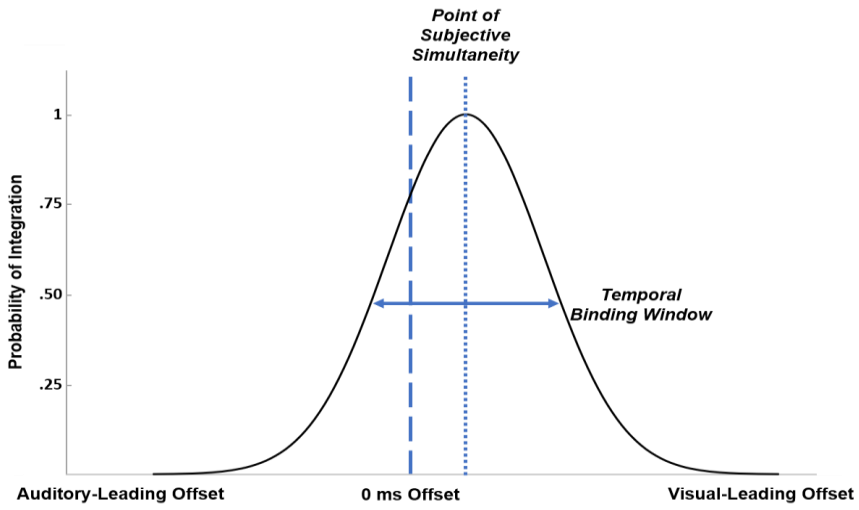


Figure 1. An example of the hypothetical outcome of an audiovisual temporal binding window measurement in which participants judge whether unimodal stimuli are or are not synchronous.

To date, there is very little literature specifically examining CI users' capacity to integrate audiovisual stimuli across temporal offsets. Generally, these limited results suggest that CI users have a similar ability to detect asynchronies as those with acoustic hearing. However, more work is needed to better understand the roles of additional variables such as participant characteristics, stimulus characteristics, and task demands in the measurement and comparison of how CI users process audiovisual asynchrony. Hay-McCutcheon *et al.* (2009) conducted a preliminary analysis of asynchrony detection in CI users in which single audiovisual words were presented across a broad range of asynchronies and participants were asked to report a simultaneity judgement (i.e. whether the presentation was synchronous or asynchronous). The results showed no significant difference between CI users and acoustic hearing participants. Additionally, they found no relationship between size of temporal binding window and intelligibility of clear, synchronous speech in CI users. Butera *et al.* (2018) evaluated whether perception of audiovisual asynchrony varies between speech and non-speech stimuli in CI users and acoustic hearing controls. Here the authors used flash/beep and syllable/viseme pairings and asked participants to judge the synchrony of the components as well as make a temporal order judgement (i.e. report whether the auditory or visual signal occurred first in the case of asynchronous presentation). Again, findings showed no significant group

differences in the range of temporal offsets over which integration occurred for either stimulus type. Notably, this experiment *did* find that CI users showed greater sensitivity in detecting asynchrony in visual-leading conditions and observed a shift in the reported point of subjective simultaneity toward smaller visual-leading asynchronies compared to hearing controls. In accounting for the latter, the authors cite the attentional principle of *prior entry* in which cues that are attended to or are highly salient are perceived as occurring before unattended or less salient stimuli (Zampini *et al.*, 2005). Given the utility of visual speech cues in resolving ambiguous speech for CI users, these signals may be perceived as especially salient and are more likely to be attended as a reliable source of speech information for this group. Thus, increased attention to visual cues could result in a perceptual bias that would shift the point of subjective simultaneity and the surrounding temporal binding window toward shorter visual-leading asynchronies without affecting its width or shape. Butera *et al.* (2018) thus concluded that CI users assign greater perceptual weight to the visual stream than the auditory stream when judging the synchrony of speech stimuli.

In these experiments, and indeed the majority of research pertaining to temporal synchrony in audiovisual integration, effect of offset on multisensory integration was measured using simple, discrete stimuli. However, temporal binding window characteristics have been shown to vary depending upon whether measurement is carried out with speech or non-speech stimuli. For example, single-syllable audiovisual speech stimuli are integrated over a wider range of offsets than are flash/beep pairs in individuals with normal hearing (Butera *et al.*, 2018; Stevenson & Wallace, 2013). Conversely, when presented with continuous audiovisual stimuli, the temporal binding window is narrower and more asymmetric for natural speech than for speech which had been rendered unrecognizable through spectral rotation temporal inversion (Maier *et al.*, 2011). Moreover, in their research with typical hearing participants, Shahin *et al.* (2017) found that degradation of the visual and auditory speech streams was associated with greater sensitivity to asynchrony, whereas when the two signals were clear, listeners were more likely perceive them as synchronous over greater degrees of offset. The authors interpreted these results as an indication that stimuli clearly perceived as audiovisual may prime the system to widen the temporal binding window. Taken together, this evidence

suggests that audiovisual perception of ongoing, naturalistic speech shows some degree of tolerance for temporal offset, the magnitude of which may depend on the SNR of the stimulus.

A final methodological consideration is that measures of audiovisual integration may be affected by cognitive biases related to task demands. When observers are asked to directly report whether unisensory signals are synchronous or asynchronous they may be primed to believe that the stimuli *should* be integrated simply because synchronous is an available option. Similarly, when presented with a temporal order judgment task in which the only possible response options are “auditory first” or “visual first”, participants may assume that the stimuli are never synchronous (Vroomen & Keetels, 2010). Indeed, many of the tasks used to assess multisensory integration are not naturalistic as they do not capture the ways in which people interact with multisensory stimuli in real-world environments.

The role of temporal synchrony in multisensory integration for CI users has yet to be studied thoroughly at the level of continuous speech intelligibility. This is especially relevant given the rapidly increasing use of online video calling and conferencing platforms becoming essential for work and education. These platforms offer some benefits over more traditional voice-only calls such as improved sound quality on some platforms and the presence of facial cues that can provide visual information that may otherwise be lost, including articulatory and emotional information. However, internet-based video calls are susceptible to a number of issues including poor video quality due to hardware or software limitations and the environments in which they are used can be noisy and visually distracting (Mantokoudis *et al.*, 2013). Of primary concern here is asynchrony between the auditory and visual speech streams which commonly arises as a result of internet connectivity issues. These asynchronies can be pervasive and significantly exceed the bounds of asynchrony that could possibly be experienced in person, both with regards to the actual extent of the temporal lag between the two streams and the scenario in which the auditory signal occurs before the visual signal. Given these limitations and the lack of research on the role of temporal asynchrony on speech intelligibility in CI users, it is not clear whether these platforms are fully accessible to this population.

1.4 Current Study

The primary aims of this study were twofold; first we examined the role of visual cues in CI users' perception of speech; and secondly, we evaluated the extent to which temporal asynchrony between the auditory and visual speech streams may affect multisensory integration and overall intelligibility for CI users compared to acoustic hearing controls. An additional goal of this research was to establish the utility of online research methodologies for the purpose of audiovisual perception research in the CI user population.

The experiments discussed here were hosted entirely online such that each participant completed all parts of the study in their own homes or other chosen environment using personal computer equipment. This closely mimics the situations and settings in which these participants might ordinarily engage in online video calling and thus offers a highly naturalistic study design. There has been a major expansion of online research in recent years and these platforms are generally found to offer a number of benefits to researchers including increased access to special populations, and relatively low costs and time invested in data collection, resulting in larger and more diverse samples (Woods *et al.*, 2015). However, perception research, perhaps particularly in the audiovisual domain, presents unique challenges for online research approaches. Because each participant uses their own computer hardware it is not possible to control stimulus-related variables such as image sizes or auditory volume and quality with the fine-grained specificity that is generally expected of psychophysics experiments conducted on specialized in-lab hardware. Despite these limitations, effective use of online research platforms may offer important advancements for conducting research with CI users in a way that does not limit potential participants due to geographical access to lab spaces, and which seeks to directly measure device outcomes in the environments and scenarios in which they are used daily.

The first component of this study provides a proof of concept that audiovisual perception research can be carried out with CI users using online research methods. The current task is adapted from the thorough in-lab investigation of McGurk illusion perception in CI users and acoustic hearing controls carried out by Stropahl *et al.* (2017). Based on their findings we were able to make hypotheses about the relative rates of

fusion for CI users compared with acoustic hearing controls. Reproducing their pattern of group-level differences in an online environment would suggest that this platform is suitable for examining multisensory effects in this population.

For the purposes of examining the role of the visual speech stream and the effect of temporal asynchrony on audiovisual speech intelligibility, we measured CI users' and acoustic hearing controls' speech comprehension accuracy in noise across a range of audiovisual asynchronies. These data allowed us to establish which temporal offsets have the greatest and least effects on intelligibility across these two groups, and reveal offsets at which 1) audiovisual integration is most facilitated; or 2) integration may fail such that the addition of the non-preferred unisensory stimulus stream actually interferes with comprehension. If asynchrony was shown to have a lesser effect on speech comprehension in CI users than controls it would suggest that CI users are more resilient to asynchrony, possibly as a result of flexibility in the temporal binding window associated with atypical audiovisual experience. Conversely, if audiovisual asynchrony impairs speech comprehension in CI users to a greater extent than controls, it may be the case that the degraded speech signals generated by a cochlear implant lead to greater dependence on the visual stream, such that CI users are inclined to attend to the visual stream even when it interferes with overall comprehension.

Chapter 2

2 Methods

This study of multisensory speech perception in CI users consisted of two experiments designed to assess: 1) multisensory syllable perception; and 2) audiovisual speech-in-noise perception. All methods and analyses were carried out in accordance with pre-registered plans hosted on the Open Science Framework at https://osf.io/tj89g and https://osf.io/2t75p, respectively. Following screening, all participants were presented with a letter of information and informed assent was provided by checking a box indicating that the individual wished to participate in the study. Participants then completed an online questionnaire hosted on Qualtrics (Seattle, Washington) consisting of general demographics questions (sex, age, education, handedness, etc.) as well as questions about language experience (what languages are used, when each was learned, and in what proportion of communication each language is used). For participants who reported using cochlear implants, additional questions related to experience of hearing difficulty/loss and restoration were presented (age of hearing loss diagnosis, cause of hearing loss, audiometric details if available, side of CI device, use/side of hearing aid, years of experience with device, brand/model of device).

Prior to starting the experiment tasks, participants were instructed to prepare their environment by dimming lights, turning off music or television in the area, closing any other computer programs, and sitting squarely in front of their computer a comfortable distance from the screen at a desk or table. Participants were presented with a clip of multi-talker babble and asked to set their computer volume to a comfortable level and not to change that level for the duration of the experiments. Participants were free to choose their preferred sound output set up including speakers, headphones, earbuds, or Bluetooth streaming directly to their implant in the case of the CI user group. All participants completed the questionnaire first, followed by Experiment 1 – Audiovisual Syllable Perception, then Experiment 2 – Audiovisual Speech-in-Noise Perception. After all parts of the study were completed, the participant was provided with a debriefing form and given the option to receive a $20 gift card via email. Ethics approval for this study was obtained from the University of Western Ontario's Non-Medical Research Ethics Board.

2.1 Experiment 1 – Audiovisual Syllable Perception

2.1.1 Participants

A total of 42 participants (15 CI users, 27 controls) were recruited to participate in this study. One control participant was excluded as an outlier according to the preregistered criteria for removal of exceptionally low sentence comprehension accuracy scores (Experiment 2), suggesting either a lack of attention to the task or atypical audiovisual function. All analyses were carried out using the remaining group of 26 control participants. CI users were recruited via social media, relevant email listservs, and newsletters distributed by cochlear implant research groups and implant support and advocacy organizations. The mean age of CI users was 59.9 years (range: 26 – 78, *SD* = 15.5, 10 females; see Appendix C for hearing health history based on questionnaire responses). A matched control sample of typical hearing participants (mean age 61.2 years, range: 21 – 86, *SD* = 17.1, 20 females) was recruited through social media and the OurBrainsCAN database. This study was adapted from an in-lab by Stropahl *et al.* (2017) in which a large effect size was observed for the effect of interest (group difference in the AV incongruent condition; $d = 2.4$; U = -3.53, $n_{CI} = 8$, $\underline{n}_{Control} = 24$). Accounting for the possibility of increased variance introduced by online testing, the aim here was to detect a more conservative effect size of d = 0.8 at an alpha = 0.05 with power = 0.8. Using an allocation ratio of 2 ($n_{Control}/n_{CI}$; acknowledging that the CI group would be more difficult to recruit), a power analysis conducted using G*Power (Erdfelder *et al.*, 1996) suggested a total sample size of 45 (15 CI users, 30 controls).

All participants were required to be at least 18 years old, fluent in English, have normal or corrected to normal vision, and no history of neurological disorders. Additionally, CI users were required to self-report having acquired, severe to profound hearing loss resulting in cochlear implant. By contrast, typical hearing control participants reported no known hearing disorder or difficulty. In order to complete the online experiment, all participants were required to have stable internet access, and a computer with a keyboard and hardware for sound output.

2.1.2 Stimuli

This experiment consisted of an online adaptation of a task designed to elicit the McGurk effect (McGurk & Macdonald, 1976) modeled on the methods described by Stropahl *et al.* (2017). All speech syllable stimuli were drawn from the OLAVS set (Stropahl *et al.,* 2017) which comprises three auditory syllables ("Ba", "Ma, "Pa") and five visual syllables ("Da", "Ga", "Ka"," Na", and "Ta") recorded by eight different native German talkers. In the interest of limiting the total run time of this experiment, a subset of four talkers (1, 3, 6, and 8) were selected such that there were two male and two female talkers spanning a broad range of reported fusion frequencies (Stropahl *et al.,* 2017). These stimuli were presented in three different conditions: audio-alone, visual-alone, and incongruent audiovisual (the 'McGurk' condition). The audiovisual combinations presented are shown in Table 1. In the visual-alone and audiovisual conditions, visual stimuli were presented at 75% of the total monitor height on a mid-grey background. In audio-alone trials, a black fixation cross appeared in the center of the screen on a mid-grey background.

Table 1: The twelve auditory and visual syllable combinations (left column) in the OLAVS set and the associated four-alternative forced choice response options (right column) that were presented to the participant.

A – V Stimulus	Four-Alternative Forced Choice Options (A, V, Fusion 1, Fusion 2)
Ba-Da	Ba, Da, Ga, Pa
Ba-Ga	Ba, Ga, Da, Ma
Ba-Ka	Ba, Ka, Ga, Da
Ba-Na	Ba, Na, Ga, Da
Ba-Ta	Ba, Ta, Pa, Da
Ma-Ga	Ma, Ga, Na, Ba
Ma-Ta	Ma, Ta, Na, La
Pa-Da	Pa, Da, Ka, Ta
Pa-Ga	Pa, Ga, Ka, Ta
Pa-Ka	Pa, Ka, Da, Ta
Pa-Na	Pa, Na, Ka, Ta
Pa-Ta	Pa, Ta, Da, Ka

Note. Table adapted from Stropahl *et al.* (2017)

2.1.3 Procedure

Each unique stimulus token and audiovisual combination across the four talkers was presented five times, resulting in a total of 60 audio-alone trials, 100 visual-alone trials, and 240 audiovisual McGurk trials. All trials were blocked by condition and each participant received the three conditions in a random order. The order of trials within each condition was pseudo-randomized such that no stimulus token was repeated on back-to-back trials.

Syllable perception was measured using a closed set, 4-alternative forced choice paradigm. Each trial consisted of the presentation of a syllable followed by a response screen in which the participant was prompted to select the correct syllable from four options using the arrow keys. The positions of the various options were pseudo-randomized across the up, down, left, and right response options. For auditory-alone and visual-alone conditions, participants were instructed to select the syllable they heard or saw, respectively, from options which included the target syllable ("Ba", "Ma, "Pa" for auditory, "Da","Ga","Ka","Na","Ta" for visual) and 3 randomly selected foils. In the case of audiovisual McGurk trials, participants were instructed to select the syllable that was *heard* making the correct response the aurally-presented syllable. The response options for this condition always included the aurally-presented syllable, the visually-presented syllable, and two syllables representing the most commonly perceived illusory percepts for the given stimulus pair as reported by MacDonald & McGurk (1976) and validated in a pilot study reported by Stropahl *et al.,* 2017; see Table 1 for summary. It should be noted that while Stropahl *et al.* (2017) label these response options as “fusion” syllables, the extent to which they reflect true phonemic fusion of the two presented syllables may vary across the set. Individual performance was measured as the percent correct syllable identification for each experimental condition.

2.1.4 Analysis

There were two measures of interest analyzed in this experiment: syllable identification accuracy across presentation modality, and McGurk trial response type.

To evaluate overall performance, the accuracy with which the two groups reported the syllables across each of the three modality conditions was compared. The

McGurk illusion tends to be perceived almost always or almost never depending on the individual and the specific stimulus (Basu Mallick *et al.*, 2015); accordingly, the response distribution seen here was non-normal and non-parametric analyses were conducted. Correct syllable identification was analyzed using three Mann-Whitney U tests which compared the proportion of all tokens that were identified correctly in each of the stimulus presentation modalities for each hearing group. In the case of audiovisual trials, the correct response was the aurally-presented syllable.

To further evaluate group-level differences in audiovisual integration specifically, responses to audiovisual trials were analyzed by type. Three additional Mann-Whitney U tests were conducted to compare the proportion of these trials in which participants reported the auditory syllable, the visual syllable, or either of the two fusion syllables across groups. All alpha levels in the non-parametric analyses described above were Bonferroni corrected for repeated measures where necessary.

2.2 Experiment 2 – Audiovisual Speech-in-Noise Perception

2.2.1 Participants

The participants in this experiment were the same as those who completed Experiment 1.

2.2.2 Stimuli

This experiment consisted of a listening and transcription task. Speech stimuli were drawn from the MAVA Corpus (Aubanel *et al.*, 2017), a list of 205 sentences selected from the original IEEE sentence set (Rothauser *et al.*, 1969) and normalized for phonetic balance. All MAVA corpus sentences were recorded by a native Australian English female talker with high quality video and audio. Each sentence had a duration of approximately three seconds and consisted of between five and ten words. Each sentence contained five keywords that were scored for comprehension. Sentences were presented in three different modality conditions: audio-alone, visual-alone, and audiovisual. In the visual-alone and audiovisual conditions, visual stimuli were presented at 75% of the total monitor height on a mid-grey background. In audio-alone conditions, a white fixation cross appeared in the center of the screen on a mid-grey background. Additionally, the temporal asynchrony of audiovisual stimulus presentation was manipulated to give nine

conditions: four audio-leading (-100, -200, -300, -400 ms), synchronous, and four visual-leading (100, 200, 300, 400 ms).

All sentences were presented in twelve-talker babble to encourage attendance to the visual speech cues (where present) and mimic a more ecologically valid listening environment. Speech-in-noise performance is significantly impaired in CI users relative to typical hearing controls (Hochberg *et al.*, 1992; Spriet *et al.*, 2007; Yang & Fu, 2005), such that there is no single signal SNR that would not be affected by floor or ceiling effects, respectively. To avoid measures of multisensory gain being uninterpretable due to such effects, stimuli were presented to each group at an SNR previously shown to result in approximately 60% performance for auditory-alone speech recognition; thus, audio-alone and multisensory stimuli were presented at an intended SNR of +9 dB for CI users and at 0 dB intended SNR for the typical hearing control group. Here, intended SNR levels refer only to the relative intensities of the target speech and background noise as programmed in the experimental platform and presumably presented by the computer hardware. This measure does not reflect any additional environmental noise which may have contributed to the participants' total experienced SNR.

2.2.3 Procedure

Sixteen sentences were presented in each of the 11 conditions (audio-alone, visual-alone, and nine audiovisual asynchrony conditions) for a total of 176 sentence trials. Accordingly, the 176 sentences from the 205-sentence MAVA Corpus that showed the highest accuracy in quiet during pilot testing were selected for the current experiment such that no sentence was presented more than once. Sentences from all conditions were presented in random order across five experimental blocks. During each trial, a sentence was presented, and participants were then prompted to type that sentence as completely and accurately as possible. Speech comprehension in each stimulus condition (audio-alone, video-alone, each AV asynchrony) was quantified as the mean percentage of keywords correctly identified.

2.2.4 Analysis

Unimodal performance was compared across groups using a mixed model ANOVA with modality (visual-alone and auditory-alone) treated as a within-subject variable and group as a between-subjects variable.

In addition to overall accuracy across modality, indices of Bimodal Effect and Multisensory Gain were calculated to quantify audiovisual integration. To examine possible interference caused by audiovisual asynchrony, we computed Bimodal Effect, defined here as the change in performance related to the availability of both auditory and visual speech information. This was calculated at each offset for each individual by subtracting their best unimodal performance level (U_{best}) from their audiovisual performance (AV) and normalizing to the amount of behavioural gain available:

$$\text{Bimodal Effect} = ((AV - U_{best})/(100 - U_{best}))$$

Here, negative values indicate temporal offsets at which audiovisual information interfered with speech comprehension to some degree.

Multisensory Gain was also computed to describe the perceptual benefit of integrating auditory and visual speech information. This was calculated at each offset for each individual by computing the difference between the observed performance at that audio-visual offset (AV) and the expected multisensory accuracy based on unisensory responses (p(A) and p(V) represent the probability of a correct response given the auditory and visual information alone, respectively):

$$\text{Multisensory Gain} = AV - (p(A) + p(V) - [p(A) \times p(V)])$$

Here, positive values indicate that the presence of audiovisual speech information had a faciliatory effect on comprehension.

For each index, a mixed model ANOVA was performed in which temporal offset (nine levels spanning -400 to +400 ms) was treated as a within-subject variable and hearing group as a between-subjects variable. For significant interactions, post hoc t-tests[1] were carried out to examine groupwise differences at each offset. The standard $p<.05$ criteria was used for the ANOVA and post hoc t-tests were Bonferroni corrected.

---

[1] Note that Games-Howell post hoc tests were specified in the preregistration however this approach was later determined to be inappropriate for the two-way ANOVAs described here.

Chapter 3

3 Results

3.1 Experiment 1 – Audiovisual Syllable Perception

Group average results for CI users and typical hearing controls in the syllable perception task are shown in Figure 2. A first set of analyses examined group differences in correct phoneme identification across presentation modalities (Figure 2a). Mann-Whitney U tests were conducted to compare syllable identification accuracy between the two groups across the audio-alone, visual-alone, and incongruent audiovisual ('McGurk') conditions. The Bonferroni corrected alpha levels for these comparisons was $p = .017$. In the audio-alone condition, typical hearing controls showed high accuracy in identifying the heard syllable ($M = 95.51\%$; $SD = 7.64\%$) whereas the CI users were significantly less accurate ($M = 79.33\%$; $SD = 19.59$; $U = 64.5$; $p < .001$, $d = 1.3$). Both groups showed poorer ability to correctly identify syllables in the visual-alone condition, with the typical hearing control group performing significantly less accurately ($M = 30.85\%$; $SD = 4.60\%$) than CI users ($M = 35.73\%$; $SD = 6.04\%$; $U = 99.5$; $p = .01$, $d = .87$). While performance was markedly decreased relative to auditory-alone performance, both groups' accuracy was above chance for visual only syllables (Control $t(25) = 6.48$, $p < .001$; CI $t(14) = 6.88$, $p < .001$). In the audiovisual incongruent ('McGurk') condition, the aurally-presented syllable was considered the 'correct' phoneme (as per the task instructions). In this condition, no significant difference between groups was observed (Control $M = 22.44\%$, $SD = 26.51\%$; CI $M = 12.67\%$, $SD = 19.53\%$, $U = 141.0$, $p = .147$, $d = .46$).

Subsequent analyses further examined group differences in the distribution of response types in McGurk trials (Figure 2b). In this audiovisual condition, the response options reflected the syllable presented via each modality and two common fusion syllables for each stimulus pair. As described above, there was no significant difference in the rate at which groups chose the aurally-presented syllable in the audiovisual condition. In addition, there was no significant difference between groups in the rate of fusion syllable responses (Control $M = 63.14\%$, $SD = 22.24\%$; CI $M = 65.97\%$, $SD =$

16.93%, *U* = 185.0, *p* = .797, *d* = .09). CI users did, however, more often select the visual syllable when compared to controls (CI *M*= 21.33%, *SD* = 10.89; Control *M* = 14.42%; *SD* = 4.57, *U* = 90.50, *p* = .005, *d* = 1.0).

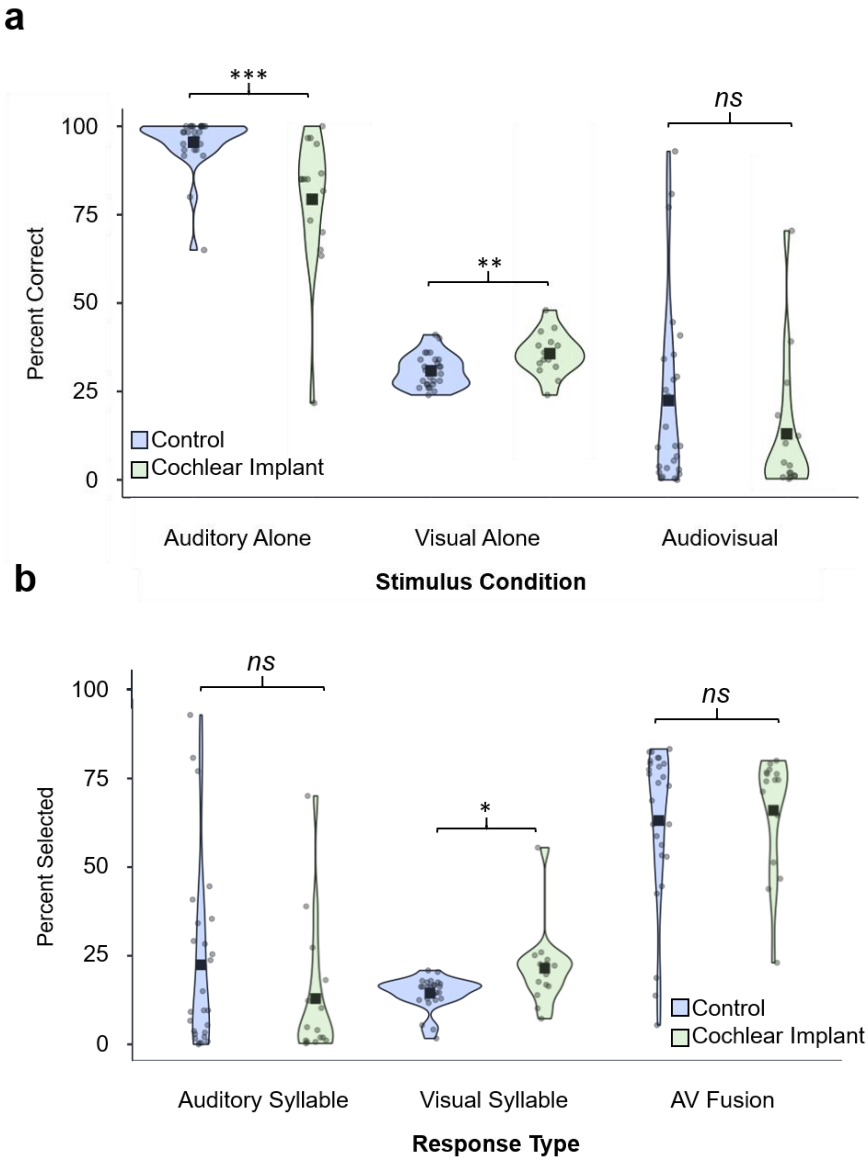


Figure 2. a) The percent correct syllable identification of the control group (blue) and CI group (green) in each of the three conditions. Per experiment instructions, the correct syllable in the incongruent audiovisual ('McGurk') condition was the auditory stimulus. b) The percent of response types selected in the McGurk condition separated by the auditory, visual, or fused syllable. Each dot represents a single participant.
* = <.05, ** = <.01, *** = <.001

In comparing these results to those reported by Stropahl *et al.* (2017), the raw accuracy scores show very similar levels of performance across participant groups between the two studies in both unisensory conditions with the exception that the current CI group performed more accurately, especially in the auditory only condition. Conversely, in the incongruent audiovisual condition, the typical hearing control participants here reported the auditory syllable less often than in the study by Stropahl *et al.* (2017), and instead reported a larger proportion of fused percepts.

3.2 Experiment 2 – Audiovisual Speech-in-Noise Perception

This experiment aimed to examine the accuracy of speech-in-noise perception in CI users and typical hearing controls and assess the effects of temporal offset on multisensory processing. Differences in unimodal performance between the groups (Figure 3) were assessed with a mixed model ANOVA with modality (visual-alone and auditory-alone) treated as a within-subject variable and group as a between-subjects variable. This reveled a significant main effect of modality with accuracy being higher for the auditory-alone condition than the visual-alone condition ($F(1, 39) = 106.46$, $p < .001$, $d = 1.8$). There was also a significant interaction between modality and hearing status ($F(1, 39) = 12.51$, $p = .001$, $d = 2.0$), whereby auditory performance was better in the control group than the CI group, while the opposite trend was observed for visual-alone stimuli. Bonferroni corrected post hoc contrasts showed that control participants performed significantly better than CI users in the auditory-alone condition ($t(39) = -2.51$, $p = .016$, $d = -.81$) but revealed no significant group difference in the visual- alone condition ($t(39) = 2.19$, $p = 0.034$, $d = .71$), where floor effects were apparent for both groups.

For audiovisual speech, the effect of temporal offset on the performance of the two groups is illustrated in Figure 4, wherein accuracy at each offset was normalized to an individual's performance in the synchronous condition (0 ms offset), and then averaged across groups.

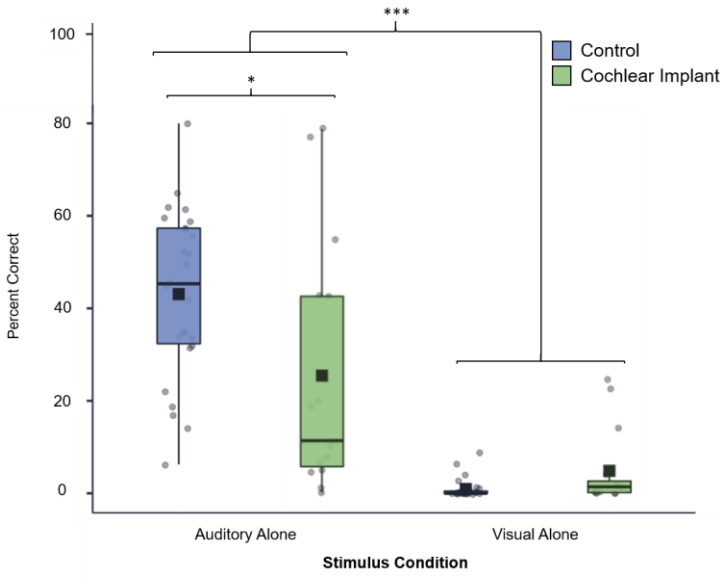


Figure 3. Accuracy of CI and control groups in unisensory conditions as percent of key words correctly reported. Each dot represents a single participant. *Note*, due to the breadth of individual scores across the auditory-alone condition and obvious floor effects present across groups in the visual-alone condition, violin plots were uninterpretable for these data. * = <.05, ** = <.01, *** = <.001

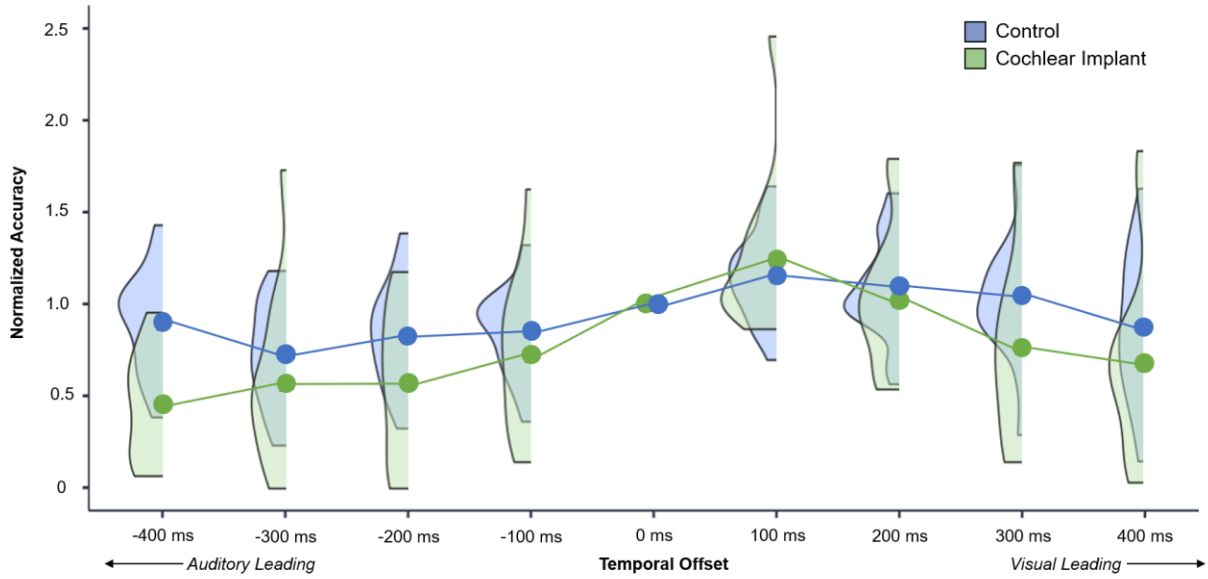


Figure 4. Overall effect of temporal offset on accuracy in CI and control groups normalized as proportion of accuracy at synchrony (0 ms offset).

Further analyses were then conducted to examine multisensory interactions between the two groups and across the range temporal offsets. To do so, accuracy data were transformed to create two indices; Bimodal Effect and Multisensory Gain. Bimodal Effect was calculated to determine whether the addition of the visual stream caused interference with overall speech perception accuracy at any offset(s). This index was calculated at each offset by subtracting an individual's best unimodal performance level from their audiovisual performance and normalizing to the amount of gain available (Bimodal effect = $((AV - U_{best})/(100 - U_{best}))$). Differences in Bimodal Effect between groups (Figure 5) were analyzed using a mixed model ANOVA with temporal offset (9 levels; -400, -300, -200, -100, 0, 100, 200, 300, and 400 ms) treated as a within-subject variable and group as a between-subjects variable. Results show a significant interaction between temporal offset and hearing status in which CI users showed a greater change from their unisensory baseline, especially near their peak performance level, than did the control group ($F(8, 312) = 1.94$, $p = .05$, $d = .29$). However, post hoc tests showed no significant group difference at any individual temporal offset (all uncorrected p values > .05, see Appendix D). There was also a significant main effect of offset, with both groups showing larger effects of bimodal stimuli around the point of synchrony than at extreme audiovisual asynchronies. ($F(8, 312) = 14.99$, $p < .001$, $d = .91$).

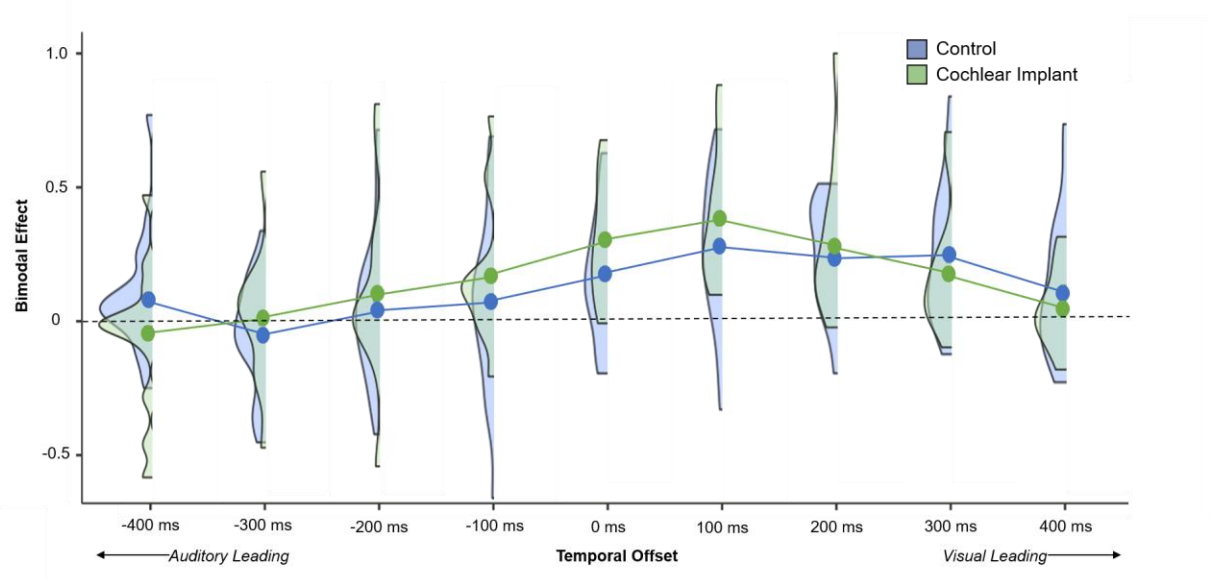


Figure 5. Index of Bimodal Effect for CI and control groups at each offset.

Finally, Multisensory Gain was calculated to describe the perceptual benefit of integrating auditory and visual speech information and resulting facilitation of overall speech perception accuracy over the range of asynchronies tested. This index was calculated at each offset for each individual by computing the difference between the observed performance at that audio-visual offset and the expected multisensory accuracy based on unisensory responses (Multisensory gain = AV – (p(A) + p(V) – [p(A) x p(V)])). In this calculation, positive values indicate that the integration of auditory and visual speech cues had a faciliatory effect on comprehension, above that which would be expected based on the combined unisensory performance levels. Differences in Multisensory Gain between groups (Figure 6) were analyzed using a mixed model ANOVA with temporal offset (9 levels; -400, -300, -200, -100, 0, 100, 200, 300, and 400 ms) treated as a within-subject variable and group as a between-subjects variable. Here too, a significant interaction between offset and hearing status was observed, with CI users exhibiting greater multisensory facilitation over a range of short duration asynchronies compared to controls ($F(8, 312) = 2.62$, $p = .009$, $d = .35$). However, post hoc tests showed no significant group difference at any individual temporal offset (all uncorrected p values > .05, see Appendix D). There was also a significant main effect of offset with the greatest facilitation related to multisensory integration taking place near the point of synchrony ($F(8, 312) = 21.87$, $p < .001$, $d = 1.1$).

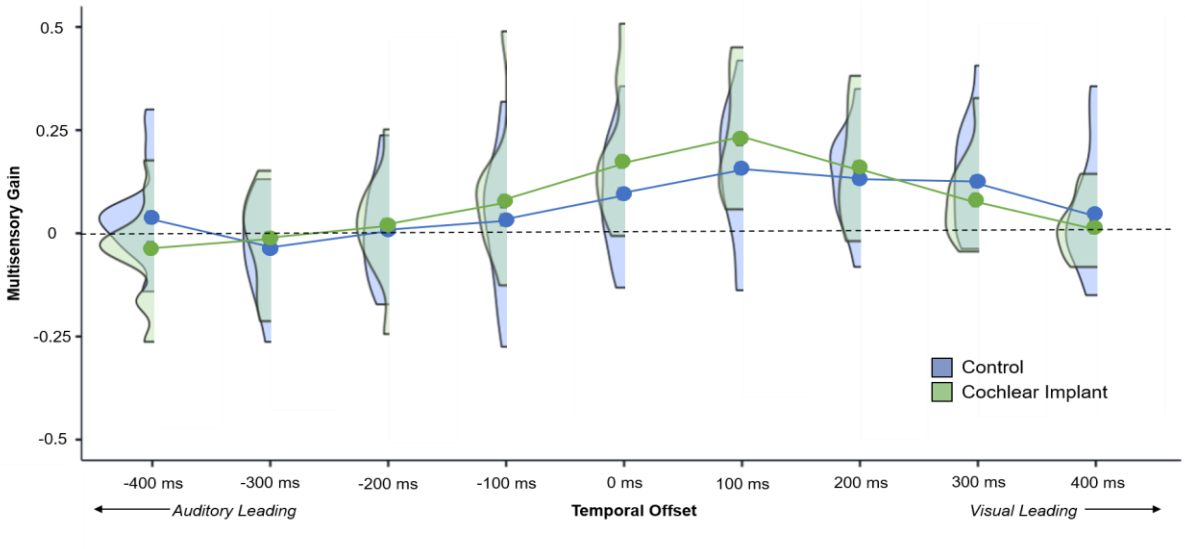


Figure 6. Index of Multisensory Gain for CI and control groups at each offset.

Chapter 4

4 Discussion

The goals of this study were two-fold: first, to examine the role that visual speech cues play in CI users' perception of audiovisual speech; and second, to evaluate the extent to which temporal asynchrony between auditory and visual speech streams affects multisensory integration and overall speech intelligibility for CI users compared to typical hearing controls. An additional goal of the study was to determine whether online data collection methods are suitable for conducting audiovisual perception research with CI users.

4.1 Experiment 1 – Audiovisual Syllable Perception

The first experiment in the current study aimed to establish the utility of online data collection for audiovisual perception research by adapting Stropahl *et al.'s* (2017) in-lab comparison of McGurk illusion perception between CI users and typical hearing controls. Participants completed the full study over the internet using their personal computers and audiovisual hardware to make judgements about stimuli presented via the Pavlovia research platform (Peirce *et al.*, 2019). Replication of Stropahl *et al.'s* (2017) in-lab study results under these conditions would provide evidence that an online approach is an effective alternative to traditional in-person data collection within this field of research.

Here, we found that both groups identified the correct phoneme more accurately in the auditory-alone condition than the visual-alone condition. However, typical hearing controls performed better than CI users in the auditory-alone condition whereas CI users outperformed controls in the visual-alone condition. Importantly, there was no difference between the two groups' ability to identify the aurally-presented syllable in the incongruent audiovisual ('McGurk') condition. Perhaps unsurprisingly, these unisensory outcomes suggest that those with typical hearing more accurately interpret auditory information while CI users are more likely to make use of the visual stream when perceiving speech. Improved visual discrimination of speech syllables may help compensate for the degraded auditory signals provided by a cochlear implant compared

to the typically developed cochlea, enabling better speech reading and subsequent improvements in speech intelligibility. Notably, that both groups reported similar rates of perceiving the aurally-presented syllable on incongruent audiovisual trials differs from Stropahl *et al.'s* (2017) finding that controls were significantly more likely than CI users to report what was heard on McGurk trials.

Looking more closely at responses to incongruent audiovisual trials, the patterns observed in the current study are qualitatively similar to those observed by Stropahl *et al.* (2017), with CI users being significantly more likely than control participants to report the visually-presented syllable while controls were more likely than CI users to report what they heard (although this latter contrast failed to reach statistical significance). However, we observed no significant difference between the two groups in the frequency of perceiving an illusory fusion syllable, while Stropahl *et al.* (2017) found CI users significantly more likely than controls to experience fusion.

4.1.1 Sensory cue weighting for optimal perception

The overall pattern of results demonstrated here speaks to a shift in perceptual bias between CI users and individuals with typical acoustic hearing. Generally, sensory organs receive signals from the environment that include some amount of ambiguity, and the goal of the associated perceptual systems is to interpret the most likely underlying nature of those signals. For audiovisual speech, the auditory and visual sensory components are typically concordant and redundant which provides an abundance of information from which systems involved in the perception of speech can resolve the signal. Ideally, over time, these systems learn about the relative reliability of representations arising from each sensory modality and determine how these representations should be most appropriately weighted in resolving speech signals across a variety of listening scenarios. According to the principle of inverse effectiveness, when unisensory stimuli are clear, multisensory integration may confer little benefit, as simply attending to unisensory cues can be sufficient for intelligibility (e.g., auditory speech perception in the absence of background noise). Conversely, when the representation of one modality is ambiguous and consequently less effective/reliable (e.g., the stream of auditory cues provided by a cochlear implant), these systems should adaptively rebalance such that less weight is

given to the degraded signal and more weight is placed on complementary signals that may improve intelligibility. By examining the patterns of responses to incongruent auditory and visual signals in Experiment 1, we were able to test the relative weightings assigned to cues arising from each modality by measuring the extent to which behavioural responses reflected visual and auditory inputs. The overrepresentation of visually-presented syllables reported by CI users compared to typical hearing controls suggests that the perceptual systems of CI users shift their bias to place greater weight on visual cues than those with typical acoustic hearing.

4.1.2 Comparison to lab-based study

Overall, the results of the current study are in alignment with the general patterns of visual bias in CI users described by existing studies (Butera *et al.*, 2018; Stropahl *et al.*, 2017). However, the effect sizes obtained here are smaller than those reported in the original Stropahl *et al.* (2017) study; where the current study observed small effect sizes (Cohen, 1988) for group differences in the rates of reporting aurally-presented (d = 0.46) or fused syllables (d = 0.09) in the McGurk condition, Stropahl *et al.* (2017) report large effect sizes of d = 2.4 and d = 2.3 for these same comparisons. Potential explanations for this discrepancy include: participant age effects, a possible effect of talker accent, and the possibility that some sensory or perceptual phenomenon unique to experiencing these stimuli in a remote, internet-based research setting affected outcomes.

With regard to age effects, perception of the McGurk illusion is known to vary with age such that older observers are more strongly influenced by the visual stream and more likely to experience fusion (McGurk & Macdonald, 1976; Sekiyama *et al.*, 2014). The CI user group and the typical hearing control group described here were age-matched (CI M = 59.9 years, control M = 61.6 years), each comprising a group of older adults. Conversely, Stropahl *et al.* (2017) acknowledged a marked age difference between their groups of CI users (M = 47 years) and controls (M = 26 years). Thus, the larger group differences that Stropahl *et al.* (2017) presented as being driven by use of a cochlear implant may in fact represent the additive effects of both hearing experience and age. The authors argue that the performance of their oldest and youngest CI users, 75 and 19 years respectively, was comparable so age effects were unlikely to be a factor. However, their

total sample size was small (8 participants) and therefore underpowered to elucidate individual differences or patterns spanning disparate age groups. To tease apart the potential effects of the online environment and participant age, further investigation is planned in which a younger sample of control participants, age-matched to the sample reported in Stropahl *et al.* (2017), will complete the online experiment described here. Replication of group differences to the full effect reported by Stropahl *et al.* (2017) would serve as evidence of the presence of an age effect compounding the extent of groupwise differences. Indeed, a preliminary analysis of data collected from university-aged participants undertaken during the piloting of this study more closely matched the patterns reported by Stropahl *et al.* (2017) (see Appendix E).

An additional difference between this study and that of Stropahl *et al.* (2017) is the native language of the participants. The OLAVS stimuli are spoken by trained, native-German speakers. Although the tokens themselves do not necessarily convey linguistic meaning, there may be subtle differences in pronunciation and enunciation of the syllables that are perceived differently by English (current study) and German (Stropahl *et al.,* 2017) speaking participants. It is possible these subtle perceptual differences may have contributed to differences in the rate of illusion perception across the two studies.

A final consideration here is whether there is something inherent to the online research environment that affects the perception of audiovisual stimuli and therefore the likelihood of experiencing the McGurk illusion. Unlike in-lab testing, participants in online research studies experience stimuli in any number of real-world environments, aspects of which may influence their perceptual experience. Nevertheless, in the incongruent audiovisual condition, CI users reported perception of a fused syllable at a rate comparable to in-lab study (65.97% of trials in the current study vs 68.62% of trials in lab [Stropahl *et al.*, 2017]), suggesting that the online platform used was very successful at eliciting multisensory integration in this population. Thus, the qualitatively similar pattern of results observed between the current study and previous in-lab work suggests that the online approach described here is a useful avenue for the study of audiovisual perception.

4.2 Experiment 2 – Audiovisual Speech-in-Noise Perception

This experiment explored the broad questions of how visual cues impact speech-in-noise perception in CI users and how audiovisual asynchrony affects multisensory processing. Here, we used continuous audiovisual sentences (MAVA Corpus; Aubanel *et al.*, 2017) presented in group-specific levels of multi-talker babble background noise. Sentences were presented in unisensory conditions and across a range of audiovisual temporal offsets to measure participants' speech intelligibility. Two potential outcomes were hypothesized: either CI users' accuracy would be *less* affected by audiovisual asynchrony than typical hearing controls due to adaptive widening of the temporal binding window; or CI users' accuracy would be *more* affected by asynchrony than controls, suggesting an increased dependence on the visual speech stream even at offsets beyond which integration and associated perceptual benefits can occur.

Here, both groups performed more accurately in the auditory-alone than visual-alone condition. Furthermore, typical hearing controls performed better than CI users in the auditory-alone condition whereas CI users outperformed controls in the visual-alone condition, though the latter did not reach significance. Indeed, group differences in auditory-alone performance are underestimated in this group-level comparison, as the two groups were presented with speech stimuli in drastically different levels of background noise (+9 dB intended SNR for implant users, 0 dB intended SNR for controls; see methods for justification). Presentation of a single SNR across groups would certainly have resulted in a much larger group difference in auditory-alone performance (a difference that would likely be uninterpretable due to floor/ceiling effects). CI users' superior performance in the visual-alone condition likely arises from their increased reliance on visual speech cues to support intelligibility both while using the implant and during the period of hearing loss which preceded implantation (see Appendix C for a detailed description of participants' hearing health history).

Compared to unisensory conditions, accuracy was enhanced for both CI users and control participants in the audiovisual condition. In both groups, peak performance occurred at an offset with a slight visual lead (approx. 100 ms), with accuracy decreasing with increasing degrees of offset. The effect of audiovisual asynchrony was asymmetrical, with a more gradual decline for visual-leading conditions and a sharper

decrease for audio-leading offsets. This asymmetry conforms to established findings in which the point of subjective simultaneity reflects natural statistics related to the relative propagation speeds of light and sound creating a greater sensitivity to audio-leading offsets (Dixon & Spitz, 1980). For both visual-leading and audio-leading offsets, controls showed more moderate decrements in accuracy with increasing offset whereas CI users showed more immediate declines in performance as offset increased. Interestingly, the control group showed a notable uptick in accuracy at the largest audio-leading offset (mean performance was greater at the 400 ms audio-leading offset than at 300 ms audio-leading; $t(25) = 3.15$, $p = 0.004$). Anecdotally, some participants reported that at this most extreme offset it was possible to establish a speech percept based on the auditory cues available, and to resolve ambiguities therein by subsequently focussing on visual cues. Thus, it is possible that the improvement in performance at this most extreme audio-leading offset reflects an advantage of audiovisual presentation, but one that is unrelated to integration of the two unisensory signals. Overall, the patterns of audiovisual accuracy observed in the current study support the idea that all listeners are affected by audiovisual asynchrony, but the extent of this impact on speech intelligibility differs as a function of hearing status.

To further examine the effect of audiovisual asynchrony on multisensory processing, raw accuracy scores were transformed into two indices of multisensory integration. First, to determine whether the addition of temporally misaligned visual speech information interfered with intelligibility at any offset, Bimodal Effect was calculated. This index compared each participant's accuracy at each offset in the audiovisual condition with their audio-alone performance (the best unisensory accuracy condition across all participants tested). A negative value at any offset would indicate that the presence of the visual stream interfered with speech intelligibility, resulting in performance below that observed for the auditory stream alone. However, bimodal effect did not dip significantly below zero for either group at any temporal offset. These data therefore suggest that the addition of the visual stream, even at large degrees of asynchrony did not significantly interfere with speech intelligibility.

A second index, Multisensory Gain, was calculated to capture the extent to which the addition of the visual stream enhanced perception, thereby facilitating intelligibility.

This index is superior to the Bimodal Effect for quantifying *gains* related to multisensory integration because it considers gains relative to the expected accuracies across both unisensory conditions such that any observed gain can more confidently be attributed to the binding of these signals. While groupwise differences did not reach significance at any specific offset, the magnitude of the difference between CI users and controls varied as a function of offset. Similar to Bimodal Effect, CI users showed greater Multisensory Gain than typical hearing controls for synchronous presentations, and for shorter duration audiovisual asynchronies. Gain declined to comparable levels across the two groups at intermediate offsets before ultimately inverting. Taken together, this pattern of results seen across offsets in both indices suggest that both CI users and typical hearing controls benefit most from the addition of visual speech information when it is in close temporal alignment with the auditory stream. Moreover, CI users derive greater benefit from the presence of a complementary visual speech stream when asynchrony between streams is minimal but see less enhancement than typical hearing controls at larger asynchronies.

4.2.1 Measuring sentence intelligibility in the real world

While thorough investigation of fine-grained perceptual processing in CI users has important implications for fundamental issues of auditory and multisensory processing and perception, a full understanding of how perceptual processing occurs in naturalistic listening scenarios is equally valuable. Many studies involving CI users attempt to control for as many sensory variables as possible by using noise-attenuating booths, high fidelity audio speakers, and other specialized equipment. Moreover, these studies are often designed for small, highly specified samples of CI users with comparable implant sidedness, duration of implantation, etc. These approaches are advantageous for delineating underlying factors that mediate group differences between CI users and individuals with typical acoustic hearing. However, with increasing control of stimulus presentation, the experience of the speech or other sensory signal becomes less comparable to what is experienced in daily life, thereby limiting the generalizability of results to varied listening scenarios and device experiences. Given that the goal of cochlear implantation is to provide improved accessibility to the auditory components of daily life, it is crucial that additional research pertaining to naturalistic listening

environments and scenarios be undertaken. Thus, the ecological validity provided by the use of naturally spoken speech stimuli presented via a participant's preferred online audiovisual environment is a major strength of this work.

Despite the practical and theoretical advantages of online research, there are limitations to this approach that need to be considered. A concern in the present study was the extent to which it is possible to control the signal to noise ratio experienced by each participant. Speech reception threshold measurements are known vary depending on characteristics of the target stimulus, the type of background noise signal, and participant characteristics. In consideration of these factors, intended SNRs of +9 dB for CI users and 0 dB for control participants were chosen to be comparable to those previously identified as resulting in approximately 60% intelligibility for audio-only speech-in-noise (Hochberg *et al.*, 1992; Spriet *et al.*, 2007; Yang & Fu, 2005). However, auditory-alone performance observed here was significantly lower than this target (25% for CI users; 41% for controls).

Due to the realities of at-home testing it is very likely that the experienced SNR was not equal across participants (nor equal to the intended SNR) which may have impacted performance in the current study. In the absence of sound-attenuating booths and foam-tipped earbuds, sounds other than the specifically programmed sentences and background noise presented by the current experiments are beyond the control of researchers. Although participants were asked to complete the study in a quiet place free of distractions, there is no way to know whether a given participant might have experienced some level of environmental noise that would ultimately contribute to the total experienced SNR they perceived. For example, the experience of completing the tasks alone in a private home office would be appreciably different than completing these same tasks in the relative quiet of the kitchen table away from others in the home but near a window outside of which construction noise is occurring (resulting in a reduction in the experienced SNR). Additional factors including the many different soundcards and speakers used as well as differences in devices, processors, and programming employed by CI users likely contributed to further variation in effective SNR.

In addition to sound level differences, variability in the timing of stimulus presentation (and thus, potential variance in audiovisual asynchrony) continues to be a

concern for online studies of multisensory perception. Whereas traditional in-lab experiments are typically completed by all participants on the same specialized, well-calibrated equipment, each participant who completed this experiment did so using their personal computing equipment. However, the Pavlovia (Peirce *et al*., 2019) platform has been extensively tested and found to be highly accurate in presentation timing, such that we can be relatively confident that stimulus onsets were reliably reproduced across participants (Bridges *et al*., 2020). Because screen refresh rates can introduce onset variability in the range of ±17 ms (Woods *et al.*, 2015), the current experiment was purposefully designed to use temporal offsets on the order of hundreds of milliseconds to minimize the potential for trial-to-trial variability to obscure experimental effects. While this level of granularity inherently limits the conclusions we can draw from these data, the reported outcomes demonstrate meaningful groupwise differences in accordance with the existing literature. For instance, the current experiment found that, across groups, performance was best when visual speech cues preceded auditory cues by 100 ms; previous lab-based research has indicated the point of subjective simultaneity for speech stimuli occurs with a visual-leading 120 ms offset (Dixon & Spitz, 1980; Vroomen & Keetels, 2010) suggesting that the calibration of this experiment is sensitive enough to be meaningful.

4.2.2 Effects of age on unimodal performance

In addition to effects related to real-world listening, there is also reason to believe that age effects may have affected the outcomes of the current study. While the current study was designed to recruit samples of participants aged 18 and over, sampling effects resulted in age-matched groups comprising older adult participants. Therefore, the intended SNRs selected and registered prior to conducting the experiment were not optimized to account for the effects of normal, age-related hearing loss. For example, in their study of age effects in audiovisual perception Zhou *et al.* (2019) report 50% accuracy thresholds for consonant-nucleus-consonant words in multi-talker babble at approximately +3 dB SNR for older adults with typical hearing and approximately +14 dB SNR for older CI users. Because the extent of multisensory integration is known to vary with the strength of the unisensory signals, this variance in auditory-only

performance may have had downstream effects on the levels of multisensory integration observed in this study. However, both of the indices of integration analyzed here (Bimodal Effect and Multisensory Gain) normalized audiovisual effects relative to individual unisensory performance, accounting for at least part of this variability.

To match and control for unisensory performance more accurately, a staircasing procedure could have been used to determine each individual participant's speech reception threshold prior to beginning the experiment. Subsequently, all experimental stimuli could be presented at an experienced SNR that would ensure equivalent unisensory accuracy and allow for comparisons to be made across matched perceptual experiences. However, at the time of study design, staircasing procedures were not supported by Pavlovia (Peirce *et al.*, 2019). Alternatively, thresholds for each participant could have been determined and implemented for subsequent testing across a multi-session protocol. However, this approach would have extended the duration and complexity of an already lengthy and complex paradigm and would likely have decreased participant recruitment and retention rates. Short of the described staircase procedure, future online studies of this type may be able to produce more accurate speech reception thresholds by targeting more specific sample groups and including more comprehensive screening regarding hearing ability to ensure similarity among selected participants.

4.2.3 The role of audiovisual experience on speech intelligibility

Taken together, and in consideration of the described limitations, the results of the current study indicate that an individual's hearing type (typical acoustic hearing or cochlear implant) is associated with the extent to which they will experience behavioural benefit from the presence of visual speech information in addition to auditory cues. Hearing type is further associated with the degree to which this multisensory enhancement is modulated by audiovisual temporal asynchrony. CI users were shown to experience greater behavioural gains than typical hearing controls when audiovisual speech was near-synchronous, suggesting that CI users derive more benefit from multisensory integration. Furthermore, the sharper decline in accuracy with increasing temporal offset observed for CI users suggests that multisensory processing is more sensitive to temporal offset in this population compared to typical hearing controls. This

may suggest that CI users depend more on the visual speech stream than individuals with typical acoustic hearing to compensate for degraded auditory input.

More than one possible explanation exists for the behavioural effects demonstrated by these data. It may be the case that CI users' fundamental audiovisual processes become especially adapted to facilitate integration of synchronous speech signals. Following hearing loss and restoration, multisensory neurons receive extended exposure to degraded auditory input; with this weakening of the auditory signal, the enhancement provided by complementary visual signals would be expected to increase. For CI users, this pattern of enhanced activity is likely necessary for understanding speech in most real-world listening environments. Because the vast majority of a user's experience involves temporally aligned signals (i.e. face to face communication in the real world), these neurons may become particularly sensitive to audiovisual speech signals in which there is little to no temporal offset between streams at the expense of responsiveness to less well-aligned signals. At the behavioural level, this neural adaptation may underlie CI users' particular proficiency in integrating synchronous audiovisual speech signals *and* reduced flexibility for processing anomalous asynchrony in audiovisual speech signals.

Alternatively, these behavioural findings may also be explained by an attentional effect as speculated by Butera *et al.* (2018). This explanation suggests that CI users perceive visual cues as relatively salient and preferentially attend to these rather than the degraded auditory stream. According to the principle of prior entry (Zampini *et al.*, 2005), cues that are highly salient or that otherwise capture attention tend to be perceived as occurring before other, less notable cues. Thus, an increase in attention paid by CI users to visual speech could shift the temporal binding window such that the point of subjective simultaneity is less visually-leading (i.e. closer to a true 0 ms offset) than those with typical acoustic hearing. Anecdotally, some participants in our CI user group described having used various attentional strategies for one or both experimental tasks, suggesting that adaptive attention effects likely exist in this group whether automatic or consciously applied. While both interpretations have theoretical merit, further research at the neural level and controlling for attention effects is needed to parse an underlying explanation for these behavioural results.

4.3 Caveats Related to Online Research

While online methodology shows promise for audiovisual perception research, a number of complications specific to the goals of this study and the capacity of the data collection platform arose during the course of this work. A major drawback of this study was that its length required a large number of video and audio files to be transmitted and loaded. The associated memory requirements and load times were prohibitive for a number of participants. Additionally, participant feedback suggests that the at-home nature of this approach may magnify perceived levels of fatigue while working through the 90-minute study. Furthermore, because it is largely impossible to know the full configuration of any participant's home computer (e.g. operating system, browser details, memory constraints, etc.), it is not always possible to predict or prevent playback issues which can render a willing participant's data set incomplete or otherwise unusable. Finally, perhaps especially in special populations including older adults, we found that extensive beta-testing is crucial for enabling participants to progress independently through a multi-part study. Thus, while the current study suggests that online methods are useful and appropriate in the context of audiovisual research, there remain specific issues that must be addressed at level of the individual project through careful design and implementation to make future research user-friendly, efficient, and ultimately successful.

4.4 Conclusions

The findings of this study indicate that CI users derive greater perceptual benefit from multisensory integration of auditory and visual speech cues and are more impacted by asynchrony between the two modalities than typical hearing controls. These results call into question the accessibility of increasingly ubiquitous online video calling and conferencing platforms which are vulnerable to audiovisual asynchrony well beyond the bounds of what is perceptually possible during in-person communication. These results and participant feedback underscore the necessity of features such as accurate live captioning to allow CI users to successfully make use of these platforms in cases where integration of visual speech cues is disrupted. In a similar vein, our findings suggest that online data collection shows promise for application in the field of audiovisual perception

research, including in the CI user population; but, specific platform-level challenges need to be carefully considered and addressed when developing an experiment.

Broadly, the results of this study align with the recent literature on multisensory integration in CI users. With the recognition that McGurk illusion perception alone is likely not a comprehensive indicator of the general capacity for audiovisual integration (Getz & Toscano, 2021; Wilbiks *et al.*, 2021), the results of Experiment 1 suggest that CI users do experience the integration necessary for audiovisual perception similarly to individuals with typical hearing, though there may be underlying differences in the relative weighting of auditory and visual information between the two groups. However, as Zhou *et al.* (2019) discuss, audiovisual integration is likely not a single-step process, but rather involves multiple components, each of which may vary according to individual characteristics such as hearing status and age. Indeed, the results of Experiment 2 indicate that the presence of asynchrony affects audiovisual processing differently in CI users than in controls. A greater sensitivity to asynchrony in CI users has been previously reported with regard to synchrony detection and temporal order judgment in simple flash-beep and phoneme-viseme stimuli (Butera *et al.*, 2018). These results extend on those findings, demonstrating that this sensitivity to asynchrony persists with continuous speech stimuli and has behavioural effects on intelligibility of speech signals.

In the case of continuous speech, the visual stream likely contributes meaningful information at multiple levels. Fine-grained articulatory gestures provide nuanced information at the level of phoneme, syllable, and word identification (Rosenblum & Saldaña, 1996). At a coarser level, the onset of mouth movement serves as an important cue that a speech signal is present. Speech itself may be unique from other audiovisual signals in that its salience and relevance facilitate a specific perceptual mode that is specialized for the detection and processing of phonetic information in speech signals (Tuomainen *et al.*, 2005). Thus, the onset of a visual speech stream, especially in a challenging listening scenario, may function to prime the brain to recalibrate processing of competing signals to identify and track a target speech signal. The introduction of asynchrony between the auditory and visual speech streams fundamentally interferes with the efficacy of the visual stream on all levels. In this experiment, both the onset cues and articulatory information from the visual stream were obscured to some extent across the

asynchronous conditions. Future research in this area in which the visual speech onset information and articulatory cues are independently manipulated is necessary for better understanding of the relative roles and potential interactions of these aspects of visual speech information in audiovisual integration.

Finally, while these data cannot speak directly to potential variation in temporal binding window characteristics between CI users and typical hearing controls, the observed patterns of groupwise difference indicate that there is a functionally relevant difference in processing of audiovisual speech in naturalistic settings between these two groups. Most broadly, these results further support the accepted principles that multisensory integration, and resulting behavioural benefits, are greatest when unisensory stimuli are weak and that temporal alignment is crucial to this integration. Given that the unique auditory experience of CI users is known to involve spectral degradation of the auditory signal itself and potential adaptation of temporal binding window characteristics, further investigation of these topics in this population is needed. Moreover, such research promises important insights for improving the development and programming of implant processors and therapeutic approaches, as well as for refinement of current principles of multisensory perceptual processing.

References

Aubanel, V., Kim, J, & Davis, C. (2017). *MAVA: MARCS Auditory-Visual Australian recordings of IEEE sentences*. https://doi.org/10.4227/139/59A4C21A896A3

Basu Mallick, D., F. Magnotti, J., & S. Beauchamp, M. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Buchman, C. A., Gifford, R. H., Haynes, D. S., Lenarz, T., O'Donoghue, G., Adunka, O., Biever, A., Briggs, R. J., Carlson, M. L., Dai, P., Driscoll, C. L., Francis, H. W., Gantz, B. J., Gurgel, R. K., Hansen, M. R., Holcomb, M., Karltorp, E., Kirtane, M., Larky, J., … Zwolan, T. (2020). Unilateral Cochlear Implants for Severe, Profound, or Moderate Sloping to Profound Bilateral Sensorineural Hearing Loss: A Systematic Review and Consensus Statements. *JAMA Otolaryngology–Head & Neck Surgery*, *146*(10), 942–953. https://doi.org/10.1001/jamaoto.2020.0998

Butera, I. M., Stevenson, R. A., Mangus, B. D., Woynaroski, T. G., Gifford, R. H., & Wallace, M. T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. *Scientific Reports*, *8*(1), 11345. https://doi.org/10.1038/s41598-018-29598-x

Cohen, J. (1988). The effect size. *Statistical power analysis for the behavioral sciences*, 77-83.

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *Journal of Neuroscience*, *36*(38), 9888–9895. https://doi.org/10.1523/JNEUROSCI.1396-16.2016

Desai, S., Stickney, G., & Zeng, F.-G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America*, *123*(1), 428–440. https://doi.org/10.1121/1.2816573

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/BF03195006

Dixon, N. F., & Spitz, L. (1980). The Detection of Auditory Visual Desynchrony. *Perception*, *9*(6), 719–721. https://doi.org/10.1068/p090719

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Fetterman, B. L., & Domico, E. H. (2002). Speech recognition in background noise of cochlear implant patients. *Otolaryngology - Head and Neck Surgery*, *126*(3), 257–263. https://doi.org/10.1067/mhn.2002.123044

Getz, L. M., & Toscano, J. C. (2021). Rethinking the McGurk effect as a perceptual illusion. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-021-02265-6

Hay-McCutcheon, M. J., Pisoni, D. B., & Hunt, K. K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *International Journal of Audiology*, *48*(6), 321–333. https://doi.org/10.1080/14992020802644871

Hochberg, I., Boothroyd, A., Weiss, M., & Hellman, S. (1992). Effects of Noise and Noise Suppression on Speech Perception by Cochlear Implant Users. *Ear and Hearing*, *13*(4), 263–271.

Lachowska, M., Pastuszka, A., Glinka, P., & Niemczyk, K. (2014). Benefits of Cochlear Implantation in Deafened Adults. *Audiology and Neurotology*, *19*(1), 40–44. https://doi.org/10.1159/000371609

Maier, J. X., Di Luca, M., & Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 245–256. https://doi.org/10.1037/a0019952

Mantokoudis, G., Dähler, C., Dubach, P., Kompis, M., Caversaccio, M. D., & Senn, P. (2013). Internet Video Telephony Allows Speech Reading by Deaf Individuals and Improves Speech Perception by Cochlear Implant Users. *PLOS ONE*, *8*(1), e54770. https://doi.org/10.1371/journal.pone.0054770

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meredith, M. A., & Stein, B. E. (1983). Interactions among Converging Sensory Inputs in the Superior Colliculus. *Science*, *221*(4608), 389–391.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural Coding and Perception* (pp. 234–277). Springer. https://doi.org/10.1007/0-387-28958-5_7

NIDCD. (2016). *Cochlear Implants*. NIDCD. https://www.nidcd.nih.gov/health/cochlear-implants

Park, E., Shipp, D. B., Chen, J. M., Nedzelski, J. M., & Lin, V. Y. W. (2011). Postlingually Deaf Adults of All Ages Derive Equal Benefits from Unilateral Multichannel Cochlear Implant. *Journal of the American Academy of Audiology*, *22*(10), 637–643. https://doi.org/10.3766/jaaa.22.10.2

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rosenblum, L. D. (2019). Audiovisual Speech Perception and the McGurk Effect. In L. D. Rosenblum, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.420

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 318–331. https://doi.org/10.1037/0096-1523.22.2.318

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex*, *17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Rothauser, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., Weistock, M., McGee, V. E., Pachl, U. P., & Voiers, W. D. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences*, *104*(17), 7295–7300. https://doi.org/10.1073/pnas.0609419104

Sargent, E. W., Herrmann, B., Hollenbeak, C. S., & Bankaitis, A. E. (2001). The Minimum Speech Test Battery in Profound Unilateral Hearing Loss. *Otology & Neurotology*, *22*(4), 480–486.

Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. *Frontiers in Psychology*, *5*, 323. https://doi.org/10.3389/fpsyg.2014.00323

Shahin, A. J., Shen, S., & Kerlin, J. R. (2017). Tolerance for audiovisual asynchrony is enhanced by the spectrotemporal fidelity of the speaker's mouth movements and

speech. *Language, Cognition and Neuroscience*, *32*(9), 1102–1118. https://doi.org/10.1080/23273798.2017.1283428

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., & Wouters, J. (2007). Speech Understanding in Background Noise with the Two-Microphone Adaptive Beamformer BEAM™ in the Nucleus Freedom™ Cochlear Implant System. *Ear and Hearing*, *28*(1), 62–72. https://doi.org/10.1097/01.aud.0000252470.54246.54

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. https://doi.org/10.1038/nrn2331

Stevenson, R. A., Sheffield, S. W., Butera, I. M., Gifford, R. H., & Wallace, M. T. (2017). Multisensory Integration in Cochlear Implant Recipients: *Ear and Hearing*, *38*(5), 521–538. https://doi.org/10.1097/AUD.0000000000000435

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261. https://doi.org/10.1007/s00221-013-3507-3

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529. https://doi.org/10.1037/a0027339

Stropahl, M., Schellhardt, S., & Debener, S. (2017). McGurk stimuli for the investigation of multisensory integration in cochlear implant users: The Oldenburg Audio Visual Speech Stimuli (OLAVS). *Psychonomic Bulletin & Review*, *24*(3), 863–872. https://doi.org/10.3758/s13423-016-1148-9

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio–visual speech perception is special. *Cognition*, *96*(1), B13–B22. https://doi.org/10.1016/j.cognition.2004.10.004

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Webb, R. L., Dowell, R. C., Clark, G. M., Pyman, B. C., Brown, A. M., Seligman, P. M., & Blamey, P. J. (n.d.). *The Multi-Channel Cochlear Implant*. 5.

WHO, W. H. O. (2018). *Addressing the rising prevalence of hearing loss*. World Health Organization. https://apps.who.int/iris/handle/10665/260336

Wilbiks, J., Strand, J., & Brown, V. A. (2021). *Speech and non-speech measures of audiovisual integration are not correlated*. PsyArXiv. https://doi.org/10.31234/osf.io/rdc9t

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: A tutorial review. *PeerJ*, *3*, e1058. https://doi.org/10.7717/peerj.1058

Yang, L.-P., & Fu, Q.-J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *The Journal of the Acoustical Society of America*, *117*(3), 1001–1004. https://doi.org/10.1121/1.1852873

Yawn, R., Hunter, J. B., Sweeney, A. D., & Bennett, M. L. (2015). Cochlear implantation: A biomechanical prosthesis for hearing loss. *F1000Prime Reports*, *7*. https://doi.org/10.12703/P7-45

Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. *Neuroscience Letters*, *381*(3), 217–222. https://doi.org/10.1016/j.neulet.2005.01.085

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences*, *102*(7), 2293–2298. https://doi.org/10.1073/pnas.0406460102

Zhou, X., Innes-Brown, H., & McKay, C. M. (2019). Audio-visual integration in cochlear implant listeners and the effect of age difference. *The Journal of the Acoustical Society of America*, *146*(6), 4144–4154. https://doi.org/10.1121/1.5134783

Appendices

Appendix A – Letter of Information and Consent Form

Version Date: 20/08/20

Western SocialScience

Psychology

Project Title: Audiovisual speech perception for online content

Principal Investigator: Blake Butler, Ph.D.,
Department of Psychology | Brain and Mind Institute
The University of Western Ontario
[redacted]

Introduction: Why are you here?
Dr. Blake Butler and his research team would like to invite you to participate in a study titled: "Audiovisual speech perception for online content". This study is voluntary, and participation involves completing an online survey, and a series of online tasks, all of which can be completed from the comfort of your home.

Background: What is the purpose of this study?
Dr. Butler and his team want to understand how auditory and visual information are combined when delivered over the internet. More specifically, this study aims to investigate the extent to which the addition of visual information enhances a listener's ability to understand speech.

Participate: If you would like to take part in the study, you will be asked to complete an online survey that will collect basic demographic data and information about your auditory and language experience. If you use a cochlear implant, this survey will ask about your hearing health history, age of implantation, degree of hearing loss, and implant type. You will then complete a series of tasks in which you will be presented with speech stimuli and asked to report what you've heard either by selecting the correct response on the screen, or by typing text into a box on the screen. In total, we anticipate the survey and tasks will take approximately 90 minutes to complete.

Voluntary Participation & Withdrawal: Your participation in this study is voluntary. You may elect not to participate at any time, including after the study has begun. You may leave the study at any time without affecting your compensation. If you no longer want to participate, or you do not want your data to be used in this research, you may contact Dr. Butler (see contact information at the first page) to request that your data and personal information be deleted.

Withdrawal from the current study is possible until group analyses have been completed. Additionally, you may request that your data be withdrawn from any future project/analysis for a period of up to 7 years.

Risks: There is some risk related to the storage of digital data; while these data are stored on secure servers, there is a chance that these servers could be breached. As participant names are not associated with digital files, the identity of any data subject to a breach would not be obtained.

Benefits: There will be no direct benefit to you by participating in this study.

Confidentiality: As part of our data collection, the online survey you are about to complete will ask you to provide your sex, and age. Your survey responses will be collected using an individualized link generated using a secure online survey platform called Qualtrics. Your individual survey link will only be identifiable using the master sheet described below. Qualtrics uses encryption technology and restricted access authorizations to protect all data collected. In addition, Western's Qualtrics server is in Ireland, where privacy standards are maintained under the European Union safe harbour framework. The data will then be exported from Qualtrics and securely stored on Western University's server. Access to these data is restricted to only those on the research team* and will be kept for a minimum of 7 years. Behavioural data will be collected via the Pavlovia online experimental platform. This platform will use anonymized participant identifiers, and data will be accessible to the research team, and the Pavlovia administrative team, but not to third party vendors. Across platforms, data are only identifiable using a master sheet which links your identify/contact information and the data you provide; this master sheet is accessible only to study team members*. De-identified data from this study will be shared on the Open Science Framework, which allows other researchers access to the de-identified data indefinitely. The shared data will not contain any information that could identify you.

*Representatives of the University of Western Ontario's Non-Medical Research Ethics Board may look at your study records at the site where these records are held, for quality assurance (to check that the information collected for the study is correct and follows proper laws and guidelines).

Database for future participation: If you would like to be contacted about future research studies for which you may be eligible, you can choose to have your identifiable information entered into "OurBrainsCAN: University of Western Ontario's Cognitive Neuroscience Research Registry" by the researchers of this study OR alternatively you can be given the web address of OurBrainsCAN where you are able to enter your information. This is a secure database of potential participants for research at Western University, which aims to enrol 50,000 volunteers over a period of 5 years. The information in this database will be stored indefinitely. The records are used only for the purpose of recruiting research participants and will not be released to any third party. When you are invited to participate future research studies, you will be given a full

description of what your involvement would entail. You are, of course, free to turn down any invitation. If, at any time, you decide that you do not want your contact information to be a part of this database, please contact ourbrains@uwo.ca to remove your information.

Costs & Compensation: You are eligible to receive a $20 gift card for completing this survey. In order to facilitate compensation, your email address will be shared with the vendor (giftcards.ca), but no information about your participation in a research study will be disclosed.

Questions about the Study:
If you have any questions about the study, please contact:

Blake Butler, PhD
Department of Psychology | Brain and Mind Institute
The University of Western Ontario
[redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Research Ethics [redacted].

Checking the box below indicates that you have read the letter of information, understand the nature of the study, and agree to take part. You acknowledge that you can quit the study at any time.

- ○ Yes, I have read the above description and agree to participate

Do you consent to receiving study compensation via email from giftcards.ca?

- ○ Yes
- ○ No

I consent to being added to the OurBrainsCAN: University of Western Ontario's Cognitive Neuroscience Research Registry to be contacted about future research studies for which I may be eligible:

- ○ I have already signed-up.
- ○ Yes, the researcher can enter my information into the database on my behalf.
- ○ Yes, please provide me with the link to join the database myself.
- ○ No, thank you

Appendix B – Ethics Approval

Western Research

Date: 28 August 2020

To: Dr. Blake Butler

Project ID: 116121

Study Title: An online study of auditory-visual speech intelligibility in cochlear-implant users

Short Title: Online Speech Perception in CI

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 04/Sept/2020

Date Approval Issued: 28/Aug/2020 19:18

REB Approval Expiry Date: 28/Aug/2021

Dear Dr. Blake Butler

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document version
Identifying Information	Implied Consent/Assent	24/Jun/2020	1.0
Debrief_revised	Debriefing Document	06/Aug/2020	2.0
Poster_NH_Revised	Recruitment Materials	29/Jul/2020	2.0
Poster_CI_Revised	Recruitment Materials	29/Jul/2020	2.0
Protocol_revised	Protocol	20/Aug/2020	3.0
Recruitment_Email_Revised	Recruitment Materials	20/Aug/2020	2.0
Qualtrics_Survey_Revised	Online Survey	20/Aug/2020	3.0
LOI_Revised	Implied Consent/Assent	20/Aug/2020	3.0

Documents Acknowledged:

Document Name	Document Type	Document Date	Document version
Screening Document	Screening Form/Questionnaire	06/Aug/2020	1.0
Notice_of_ineligibility	Tracked Changes Document	20/Aug/2020	1.0

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Katelyn Harris, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix C – Overview of Cochlear Implant Users' Hearing Health History

Participant	Sex, Age	Age of Hearing Loss Diagnosis	Cause of Hearing Loss	Hearing status (unaided; self-report)	*Left Ear*		*Right Ear*	
					Device	**Years of Experience**	**Device**	**Years of Experience**
1	F, 26	3	enlarged vestibular aqueduct syndrome	Profound bilateral loss	Cochlear Nucleus 7 processor; Cochlear N22 array	19	None	
2	F, 57	20	unknown	Profound bilateral loss	None		Advanced Bionics Naida Q90	18
3	F, 62	7	high fever	Profound bilateral loss	Cochlear brand CI (model unknown)	2	None	
4	F, 53	7	genetic	Profound bilateral loss	None		Advanced Bionics (model unknown)	46
5	M, 28	4	unknown	Severe to profound bilateral loss	None		Cochlear N6	21
6	F, 74	51	believed to be hereditary	Profound loss; Moderate sloping to profound (L)	Phonak Naida V90-UP	23	Cochlear Nucleus N7 CP 1000 processor; Cochlear Profile C1512 array	2
7	F, 70	39	genetic	Severe to profound bilateral loss	Cochlear Nucleus 7	5	Resound (Hearing Aid)	25
8	M, 57	53	Meniere's Disease	Severe to profound bilateral loss	Oticon OPN S miniRITE (Hearing Aid)	3	Med-el Rondo 2 processor; Med-el Synchro NY array	1

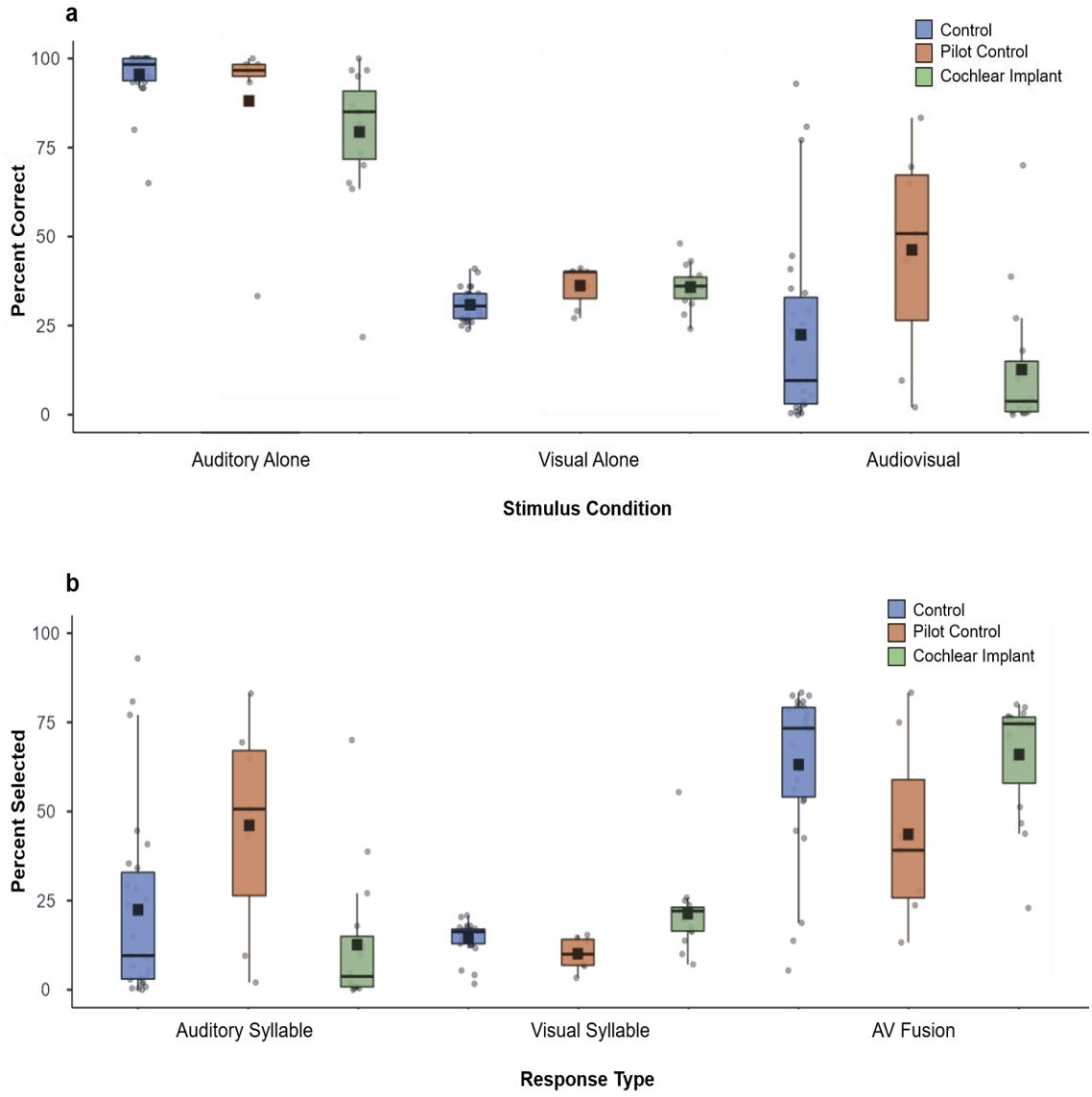
9	F, 74	40	unknown	Severe to profound bilateral loss	ReSound (Hearing Aid)	18	Cochlear Nucleus 7	2
10	M, 67	30	scarring of ear drum	Severe to profound bilateral loss	Cochlear CI612	10	Cochlear CI532	3
11	F, 60	6	ear infections	Profound bilateral loss	Cochlear Kanso 2	0	Cochlear Kanso	4
12	M, 78	42	unknown	Profound bilateral loss	Med-el Sonata	9	Med-el Sonata	13
13	F, 57	13	Oste-sclerosis	Profound bilateral loss	Med-El Sonnet	6	Med-el Sonnet	6
14	F, 75	3	unknown	Profound bilateral loss	Cochlear N6	21	Cochlear N6	16
15	M, 61	8	congenital	Severe to profound bilateral loss	Cochlear Kanso 2	0	Cochlear Kanso 2	2

Appendix D – Post Hoc Comparisons for Experiment 2, Bimodal Effect and Multisensory Gain Indices

	-400 ms	**-300 ms**	**-200 ms**	**-100 ms**	**0 ms**	**100 ms**	**200 ms**	**300 ms**	**400 ms**
Bimodal Effect	t = -1.74	t = 0.95	t = 0.58	t = 0.96	t = 1.53	t = 1.1	t = 0.26	t = -0.99	t = -0.94
	p = 0.09	p = 0.349	p = 0.563	p = 0.341	p = 0.134	p = 0.276	p = 0.796	p = 0.33	p = 0.354
Multisensory Gain	t = -2.16	t = 0.61	t = 0.11	t = 0.93	t = 1.74	t = 1.51	t = 0.71	t = -1.29	t = -0.78
	p = 0.037	p = 0.544	p = 0.91	p = 0.357	p = 0.09	p = 0.138	p = 0.484	p = 0.204	p = 0.439

Note. All dfs = 39

Appendix E – Preliminary Plots of Pilot Data from University-Aged Control Participants



Pilot data for university-aged participants (n = 7, 5 females, age M = 24.9 years, range: 20 – 36 , SD = 6.5, shown in orange) compared to the reported CI user (green) and typical hearing control (blue) groups **a)** The percent correct syllable identification of the pilot control group (orange) and CI group (green) in each of the three conditions. Per experiment instructions, the correct syllable in the incongruent audiovisual ('McGurk') condition was the auditory stimulus. **b)** The percent of response types selected in the McGurk condition separated by the auditory, visual, or fusion token component. Each dot indicates a single participant.

Curriculum Vitae

Name: Cailey Salagovic

Post-secondary Education and Degrees:

University of Western Ontario
London, Ontario, Canada
2019-present (to be completed 2021) M.Sc. in Psychology
Thesis title: *Audiovisual integration in cochlear implant users and typical hearing controls: A study of group differences in syllable perception and effect of asynchrony on speech intelligibility*

University of Colorado Denver
Denver, Colorado, United States
2014-2018 B.S. (Hons) in Psychology
Thesis title: *Crossmodal effects of irrelevant auditory stimuli on saccadic eye movement behavior*

University of Denver
Denver, Colorado, United States
2009-2013 B.A. in Music

Honours and Awards:

Western Social Science Graduate Research Award, 2020-2021
Undergraduate Research Opportunity Program Grant, 2017-2018
Samuel Priest Rose Memorial Scholarship, 2017-2018
University of Denver Provost Scholarship, 2008-2013
University of Denver Music Merit Scholarship, 2009-2013

Related Work Experience

Teaching Assistant
University of Western Ontario
Introduction to Research Methods, 2020-2021
Human Sexuality, 2019-2020

Professional Research Assistant
University of Colorado Denver
Laboratory for Integrative Vision 2018-2019

Teaching Assistant
University of Colorado Denver
Introduction to Psychology, 2016-2018

Publications:

Salagovic, C. A., & Leonard, C. J. (2021). A nonspatial sound modulates processing of visual distractors in a flanker task. *Attention, Perception, & Psychophysics, 83*(2), 800-809.

Supervisory and Presentation History
Papers in peer-reviewed journals (*N* = 1)
First-author conference presentations (*N* = 5)
Oral Presentations (*N* = 4)
Student supervision undergraduate thesis (*N* = 1), University of Western Ontario.

Conference Presentations

Salagovic, C. A., Stevenson R. A., & Butler, B. E. (March, 2021). 'Sorry, my internet is spotty': how audiovisual lag affects speech comprehension in cochlear implant users. *Poster* at the Cognitive Neuroscience Society Meeting. International - *virtual*.

Salagovic, C. A., Stevenson R. A., & Butler, B. E. (March, 2021). 'Sorry, my internet is spotty': how audiovisual lag affects speech comprehension in cochlear implant users and those with typical hearing. Western Research Forum. *Oral Presentation* at University of Western Ontario – *virtual*.

Salagovic, C. A. (November, 2020). Challenges of Moving Research Online Panel. *Panelist* at the Kids Brain Health Network Conference. Canada - *virtual*.

Salagovic, C. A. & Leonard, C. J. (November, 2018). Auditory Alerting Modulates Processing of Peripheral Visual Stimuli in Eriksen Flanker Task. *Poster* at the Auditory Perception, Cognition, and Action Meeting. New Orleans, LA.

Salagovic, C. A. & Leonard, C. J. (April, 2018). The Effect of Irrelevant Sounds on Eye Movements During Visual Search. *Oral Presentation* at University of Colorado Denver Research and Creative Activities Symposium. Denver, CO.

Salagovic, C. A. & Leonard, C. J. (February, 2018). Crossmodal effects of irrelevant auditory stimuli on saccadic eye movement behavior. *Oral Presentation* at University of Colorado Denver Psychology Department Colloquium – Undergraduate Data Blitz. Denver, CO.

Salagovic, C. A. & Leonard, C. J. (November, 2017). Crossmodal effects of irrelevant auditory stimuli on saccadic eye movement behavior. *Poster* at Auditory Perception, Cognition, and Action Meeting. Vancouver, British Colombia.

Salagovic, C. A. & Leonard, C. J. (April, 2017). What We Hear and Where We Look: effects of irrelevant auditory stimuli on eye movement behaviors. *Poster* at Rocky Mountain Psychological Association. Denver, CO.

Salagovic, C. A. & Leonard, C. J. (April, 2017). Listening with the Eyes: the effect of auditory stimuli on visual attention. *Poster* at University of Colorado Denver Research and Creative Activities Symposium. Denver, CO.