Audiovisual Integration During Novel Word Learning Among School-Aged Children with Cochlear Implants

Kristen Elizabeth Thompson Thornton

University of Tennessee Health Science Center

Follow this and additional works at: https://dc.uthsc.edu/dissertations

Part of the Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons, Speech and Hearing Science Commons, Speech Pathology and Audiology Commons, and the Therapeutics Commons

Recommended Citation
Audiovisual Integration During Novel Word Learning Among School-Aged Children with Cochlear Implants

Abstract

Objective. It is well established that being able to see someone’s mouth move as they speak boosts speech perception for children with cochlear implants (CIs). Thus, children with CIs are often instructed to orient themselves toward the person they are listening to, to gain access to visual speech cues. Children with CIs who are better “audiovisual integrators,” or those who experience an auditory-visual (AV) enhancement effect (higher performance for AV information than auditory-alone (AO) or visual-alone (VO)), are more likely to have better speech and language outcomes after receiving their CI than children with poorer AV integration skills. While AV integration of speech appears to be intimately tied with speech perception as well as speech and language development, its role in vocabulary acquisition is not well understood. This study examined novel word learning across two tasks, AV and AO, and sought to answer the following questions: (1) How does access to AV information impact novel word learning success for children with CIs and children with normal hearing (NH) listening to normal and CI-simulated speech? (2) How do individual patterns of visual attention during learning relate to individual word learning outcomes? (3) What measured factors (hearing history, device characteristics, maternal education level, etc.) contribute to novel word learning across AV and AO tasks? Methods. Twelve children with CIs (M = 7 years; 9 months) and twenty-four age- and sex-matched children with NH (M = 7 years; 8.6 months) completed two novel word learning tasks, AV and AO. Across both tasks, a female speaker was positioned on the top half of the screen and narrated a story. The corresponding story page and object to-be-learned was displayed on the bottom half of the screen. During the AO task, a black box was positioned over the speaker’s face to block access to visual speech cues. Twelve object-label pairs were presented across three blocks and word learning was assessed with a four-alternative forced-choice (4AFC) task following each block of presentations.

Results. Across listener groups, children did not learn significantly more words in the AV task as compared to the AO task. Within the group of children with CIs, two subgroups of performers were noted, “higher” and “poorer” word learners. Individual visual attention patterns corresponded with individual word learning outcomes for children who use CIs in these two performance groups. Children with CIs who spent more time looking at the speaker’s mouth learned more words than children who spent less time attending to the speaker’s mouth. Additionally, earlier age of amplification was significantly correlated with better learning outcomes. Many subscales across the LEAF, a parental report of executive functioning skills, were significantly correlated with word learning in the AO and AV tasks. Outcomes on the TONI-4, a nonverbal intelligence measure, and the Blending subtest of the CTOPP-2, an assessment of phonological processing, were also correlated with learning outcomes.

Conclusions. This study found no significant main effect of task type, which suggests that encouraging children with CIs to orient themselves to the speaker they are attending to may not be sufficient to support or improve vocabulary acquisition, particularly for children who demonstrate difficulty acquiring new words. Age of amplification, age of implantation, and phonological processing skills differentiated the two performance groups. Group differences also emerged where poorer and better CI performers showed differences in their visual attention to the task. These outcomes indicate that early learning and development of strategies for word learning warrants further investigation.

Document Type

Dissertation
Degree Name
Doctor of Philosophy (PhD)

Program
Speech and Hearing Science

Research Advisor
Mark S. Hedrick, PhD

Keywords
audio-visual, cochlear implants, vocabulary, word learning

Subject Categories
Analytical, Diagnostic and Therapeutic Techniques and Equipment | Communication Sciences and Disorders | Medicine and Health Sciences | Other Analytical, Diagnostic and Therapeutic Techniques and Equipment | Speech and Hearing Science | Speech Pathology and Audiology | Therapeutics

This dissertation is available at UTHSC Digital Commons: https://dc.uthsc.edu/dissertations/544
Audiovisual Integration During Novel Word Learning Among School-Aged Children with Cochlear Implants

Author: Kristen Elizabeth Thompson Thornton

Advisor: Mark S. Hedrick, PhD

A Dissertation Presented for The Graduate Studies Council of The University of Tennessee Health Science Center in Partial Fulfillment of the Requirements for the Doctor of Philosophy degree from The University of Tennessee in Speech and Hearing Science: Audiology College of Graduate Health Sciences

November 2020
DEDICATION

For David, Bell, Kopi, and Gabby.
ACKNOWLEDGEMENTS

I would first like to thank my adviser, Dr. Mark Hedrick for his guidance and input as I completed this project. In addition, I would like to thank Dr. Jillian McCarthy for her constant support and mentorship both in and outside of the lab. I would also like to thank my committee members, Drs. Katie Faulkner, Kevin Reilly, Susie Robertson, and Krystal Werfel for providing their respective expertise and guidance. I would like to give a special acknowledgement to Dr. Patti Johnstone for allowing me to work in her lab throughout my graduate studies. The knowledge and skills I gained have proved invaluable.

I am forever indebted to my fellow doctoral students for always providing comedic relief and perspective. I would specifically like to thank Mary Easterday and Amanda Simmons for all of the venting sessions, laughter, and dirty jokes.

Last, but certainly not least, I am grateful for the support of my family. To my husband David for his never-ending love and understanding, and to my mom, who pushed me to keep going every time I wanted to quit.
ABSTRACT

Objective. It is well established that being able to see someone’s mouth move as they speak boosts speech perception for children with cochlear implants (CIs). Thus, children with CIs are often instructed to orient themselves toward the person they are listening to, to gain access to visual speech cues. Children with CIs who are better “audiovisual integrators,” or those who experience an auditory-visual (AV) enhancement effect (higher performance for AV information than auditory-alone (AO) or visual-alone (VO)), are more likely to have better speech and language outcomes after receiving their CI than children with poorer AV integration skills. While AV integration of speech appears to be intimately tied with speech perception as well as speech and language development, its role in vocabulary acquisition is not well understood. This study examined novel word learning across two tasks, AV and AO, and sought to answer the following questions: (1) How does access to AV information impact novel word learning success for children with CIs and children with normal hearing (NH) listening to normal and CI-simulated speech? (2) How do individual patterns of visual attention during learning relate to individual word learning outcomes? (3) What measured factors (hearing history, device characteristics, maternal education level, etc.) contribute to novel word learning across AV and AO tasks?

Methods. Twelve children with CIs (M = 7 years; 9 months) and twenty-four age- and sex-matched children with NH (M = 7 years; 8.6 months) contributed data from two novel word learning tasks, AV and AO. Across both tasks, a female speaker was positioned on the top half of the screen and narrated a story. The corresponding story page and object to-be-learned was displayed on the bottom half of the screen. During the AO task, a black box was positioned over the speaker’s face to block access to visual speech cues. Twelve object-label pairs were presented across three blocks and word learning was assessed with a four-alternative forced-choice (4AFC) task following each block of presentations.

Results. Across listener groups, children did not learn significantly more words in the AV task as compared to the AO task. Within the group of children with CIs, two subgroups of performers were noted, “higher” and “poorer” word learners. Individual visual attention patterns corresponded with individual word learning outcomes for children who use CIs in these two performance groups. Children with CIs who spent more time looking at the speaker’s mouth learned more words than children who spent less time attending to the speaker’s mouth. Additionally, earlier age of amplification was significantly correlated with better learning outcomes. Many subscales across the LEAF, a parental report of executive functioning skills were significantly correlated with word learning in the AO and AV tasks. Outcomes on the TONI-4, a nonverbal intelligence measure, and the Blending subtest of the CTOPP-2, an assessment of phonological processing, were also correlated with learning outcomes.

Conclusions. This study found no significant main effect of task type, which suggests that encouraging children with CIs to orient themselves to the speaker they are attending
to may not be sufficient to improve vocabulary acquisition, particularly for children who demonstrate difficulty acquiring new words. Age of amplification, age of implantation, and phonological processing skills differentiated the two CI performance groups. Group differences also emerged where poorer and better CI performers showed differences in their visual attention to the task. These outcomes indicate that early learning and development of strategies for word learning warrants further investigation.
# TABLE OF CONTENTS

## CHAPTER 1. INTRODUCTION

Overview .................................................................................................................. 1
Purpose of the Study ................................................................................................. 2
Theoretical Application ............................................................................................. 2
  - Cognitive Load Theory ....................................................................................... 3
  - The Intersensory Redundancy Hypothesis ....................................................... 3
Overview of Objectives .......................................................................................... 4
Research Questions and Hypotheses ....................................................................... 5
  - Research Question One and Hypothesis ......................................................... 5
  - Research Question Two and Hypotheses ....................................................... 5
  - Research Question Three and Hypotheses ................................................... 6

## CHAPTER 2. LITERATURE REVIEW

Speech Is a Multisensory Signal ........................................................................... 7
Auditory and Visual Speech Cues Work Together ................................................. 7
McGurk Task ........................................................................................................... 9
Multisensory Speech Perception Across Development ...................................... 9
  - U-shaped Developmental Trajectory ................................................................ 11
  - Auditory-Visual Integration in Children with Cochlear Implants ................. 13
  - Cross-Modal Reorganization .......................................................................... 13
  - Unisensory and Multisensory Perspectives ................................................... 13
  - Spectral Degradation ....................................................................................... 14
  - Auditory-Visual Speech Perception and Higher Order Cognitive Skills ...... 14
  - Word Learning Task ....................................................................................... 16
  - Vocabulary Skills of Children with Cochlear Implants ................................ 17
  - Auditory-Visual Word Learning Among Children with CIs ......................... 18
  - The Current Study .......................................................................................... 19

## CHAPTER 3. METHODODOLOGY

Overview .................................................................................................................. 20
Participants .............................................................................................................. 20
  - Participant Recruitment .................................................................................... 20
  - Children with Cochlear Implants .................................................................... 22
  - Children with Normal Hearing ........................................................................ 22
  - Inclusion and Exclusion Criteria ...................................................................... 22
  - Inclusion criteria for children with cochlear implants .................................... 22
  - Inclusion criteria for children with normal hearing ......................................... 22
  - Exclusion criteria for all children .................................................................... 22
Stimuli ..................................................................................................................... 24
  - Familiarization Stimuli .................................................................................... 24
  - Familiarization Feedback ................................................................................ 24
  - Familiar Speech Stimuli .................................................................................. 24
  - Learning Phase Stimuli .................................................................................... 26
CHAPTER 4. RESULTS

How Does Access to AV Information Impact Novel Word Learning Success for Children with CIs and Children with NH Listening to Non-Simulated and CI-Simulated Speech? .......................................................... 39

Word Learning Across Blocks ................................................................................... 40
Item Analyses ............................................................................................................ 40

How Do Patterns of Visual Attention During Learning Relate to Individual Word Learning Outcomes? ............................................................................................................. 43

Areas of Interest ........................................................................................................ 43
Individual Differences in the CI Group ..................................................................... 44
Participant Feedback Rate ........................................................................................ 51

What Factors Contribute to Novel Word Learning Success Across AV and AO Tasks? ...................................................................................................................... 51

Variables .................................................................................................................. 51
Hearing History, Device, Aural (Re)Habilitation Variables – CI Only ...................... 54
LIST OF TABLES

Table 3-1. Demographic and hearing history information for CI participants ..........23
Table 3-2. Example familiar object-label pairs ..................................................25
Table 3-3. Novel object-label pairs and corresponding lexical and visual
characteristics ....................................................................................................28
Table 4-1. Correlation between demographic variables and AV learning outcomes
across participant groups ..................................................................................63
Table 4-2. Correlation between demographic variables and AO learning outcomes
across participant groups ..................................................................................64
LIST OF FIGURES

Figure 3-1. Overview of each phase across AV and AO presentations.......................21
Figure 3-2. Example calibration procedure, the first step of the Learning phase...........31
Figure 3-3. Example of the feedback process across one AO Test trial.......................33
Figure 3-4. Example timeline of the first part of a Learning trial, prior to the onset of auditory cues ..............................................................................................................37
Figure 4-1. Mean number of words learned across AV and AO conditions for each participant group .................................................................................................................41
Figure 4-2. Mean number of words learned across AV and AO conditions for each participant, by block.................................................................................................................42
Figure 4-3. Proportion of time spent looking to each AOI across blocks for each participant group ..................................................................................................................45
Figure 4-4. Correlation between proportion of time spent looking to each AOI and proportion of words correct across Block 1 for CI participants.........................46
Figure 4-5. Correlation between the proportion of time spent looking to each AOI and proportion of words correct across Block 2 for CI participants ..........47
Figure 4-6. Correlation between the proportion of time spent looking to each AOI and proportion of words correct across Block 3 for CI participants ..........48
Figure 4-7. Correlation among number of words learned on the AV and AO tasks for children in the CI group............................................................................................................49
Figure 4-8. Correlations between the proportion of time spent looking to the speaker’s mouth and number of words learned in the AV task .........................50
Figure 4-9. Feedback rate across CI participants for each Block in the AV task.........52
Figure 4-10. Feedback rate across CI participants for each Block in the AO task........53
Figure 4-11. Correlation between word learning outcomes on the AO task and age of amplification..................................................................................................................55
Figure 4-12. Correlation between word learning outcome on the AO task and scaled score on the CTOPP-2, Blending Subtest .................................................................55
Figure 4-13. Correlation between word learning outcome on the AO task and summed score on the LEAF Attention subscale .........................................................56
Figure 4-14. Correlation between word learning outcome on the AO task and summed score on the LEAF Sustained Sequential Processing subscale

Figure 4-15. Correlation between word learning outcome on the AO task and summed score on the LEAF Working Memory subscale

Figure 4-16. Correlation between word learning outcome on the AO task and summed score on the LEAF Novel Problem Solving subscale

Figure 4-17. Correlation between word learning outcome on the AO task and summed score on the LEAF Mathematics Skills subscale

Figure 4-18. Correlation between word learning outcome on the AV task and TONI index score

Figure 4-19. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on LEAF Factual Memory subscale

Figure 4-20. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on LEAF Processing Speed subscale

Figure 4-21. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on the LEAF Visual-Spatial Organization subscale
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AO</td>
<td>Auditory-only</td>
</tr>
<tr>
<td>AOIs</td>
<td>Areas of interest</td>
</tr>
<tr>
<td>AV</td>
<td>Audiovisual</td>
</tr>
<tr>
<td>CIs</td>
<td>Cochlear implants</td>
</tr>
<tr>
<td>CMR</td>
<td>Cross-modal reorganization</td>
</tr>
<tr>
<td>CTOPP-2</td>
<td>Comprehensive Test of Phonological Processing – 2(^{nd}) Edition</td>
</tr>
<tr>
<td>CVC</td>
<td>Consonant-vowel-consonant</td>
</tr>
<tr>
<td>IRH</td>
<td>Intersensory Redundancy Hypothesis</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-stimulus interval</td>
</tr>
<tr>
<td>LEAF</td>
<td>Learning, Executive, and Attention Functioning Scale</td>
</tr>
<tr>
<td>MLST-C</td>
<td>Multimodal Lexical Sentence Test for Children</td>
</tr>
<tr>
<td>NH</td>
<td>Normal hearing</td>
</tr>
<tr>
<td>PB-K</td>
<td>Phonetically Balanced Kindergarten word lists</td>
</tr>
<tr>
<td>PTA</td>
<td>Pure tone average</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>Peabody Picture Vocabulary Test – 4(^{th}) Edition</td>
</tr>
<tr>
<td>TONI-3</td>
<td>Test of Nonverbal Intelligence – 3(^{rd}) Edition</td>
</tr>
<tr>
<td>VO</td>
<td>Visual-only</td>
</tr>
<tr>
<td>4AFC</td>
<td>Four-alternative forced-choice</td>
</tr>
</tbody>
</table>
CHAPTER 1. INTRODUCTION

Overview

Although speech perception is typically considered an auditory-only (AO) percept, what we hear is profoundly influenced by what we see. Speech perception is inherently multimodal, whereby the majority of everyday spoken communication involves the real-time integration of information from both auditory and visual modalities. Oral motor movements and acoustic output are statistically related and change in a predictable way as the articulators, like the mouth and lips, move in accordance with the changing acoustic output (Chandrasekaran et al., 2009; Yehia et al., 1998). This statistical relationship across the visual and auditory components of speech offers sensory redundancy and benefits listeners in quiet as well as “high noise and/or low signal” environments (Helfer & Freyman, 2005; Schwartz et al., 2004; Sumby & Pollack, 1954).

It is well established that the benefit of combined AV speech information extends to hearing-impaired listeners. People who use cochlear implants (CIs) often demonstrate an additive or superadditive benefit during speech perception tasks when auditory-visual (AV) cues are presented relative to AO or visual-only (VO) cues, although the degree of benefit varies across individual listeners (Bergeson et al., 2003). While much of the published data show AV benefit for speech recognition, relatively little is known about the contribution of visual speech to learning.

Children with CIs are often included in mainstream classrooms and expected to learn and perform like their normal hearing (NH) peers, but the influence of access to visual speech information for learning is not well understood. Vocabulary acquisition, or novel word learning, presents different task demands than a word or sentence recognition task. Vocabulary acquisition requires identification of a novel word, associating that word with its referent, and retaining the word-referent pair in memory. Since vocabulary acquisition requires creation of new word knowledge, it may present greater task demands than word or sentence recognition, which requires recall of known words. As such, children with CIs may utilize visual speech cues differently during word learning, as compared to speech recognition, due to the cognitive demand differences across these tasks.

For children who use CIs, utilization of speech cues which may bootstrap vocabulary acquisition is important as children with CIs fall behind their NH peers on measures of receptive and expressive vocabulary (Geers et al., 2009; Hayes et al., 2009; Geers et al., 2003) and additionally, learn new words at a slower rate (Geers et al., 2009; Geers et al., 2016; Lund, 2016). Bergeson and colleagues (2005) analyzed a dataset of 80 children with CIs who participated in a longitudinal study that measured speech perception and language skills prior to implantation and at multiple time points post-implantation, under AV, AO, and VO conditions. Outcomes were analyzed as a function of communication mode (oral communication or total communication) and age at implantation (early, before 53 months or age, or late, after 53 months of age). An AV
benefit was demonstrated for word and sentence recognition across participant groups when compared to AO and VO conditions. Children who were late-implanted (after 53 months) showed a preference for VO word and sentence materials as compared to AO, but still demonstrated high speech recognition performance in the AV condition, comparable to children who were early-implanted. Children who used the total communication method demonstrated overall poorer word and sentence recognition performance than children who used oral communication across conditions (AV, AO, VO). The authors also noted a correlation between AV speech recognition performance and performance on tasks used to measure speech intelligibility and language skills two years post-implantation where better AV speech perception was correlated with better speech and language skills measured later in development (Bergeson et al., 2005). Specifically, the utilization of AV speech cues may be indicative of overall language outcomes. Thus, understanding the role of AV integration in word learning for children with CIs could impact our understanding of the poor vocabulary outcomes observed in this group (e.g., Geers, et al., 2009; Lund, 2016).

**Purpose of the Study**

This study examined novel word learning across AV and AO tasks. The primary objective of this study was to understand the extent to which access to AV speech information impacts novel word learning success in children with CIs and children with NH. The secondary objective of this study was to characterize individual looking patterns as they relate to individual word learning outcomes for children with CIs.

**Theoretical Application**

Listeners with CIs must make sense of highly degraded acoustic-phonetic information to understand speech because CIs convey a rudimentary representation of the acoustic speech signal (Shannon et al., 1995). As a result, understanding speech is often more effortful for CI listeners, even while maintaining high levels of accuracy (Winn et al., 2015). Although speech recognition is a relatively simple task where listeners only need to identify a word or word sequence, underspecified acoustic input strains cognitive resources during perception (Pals et al., 2013; Winn et al., 2015). In contrast, during novel word learning, children must perceive and segment speech, encode word meaning, and store this information in long-term memory (Werker & Stager, 2000). Thus, if children with CIs are already pushing the limits of their cognitive capacity to understand the degraded speech input relayed from the CI, then they likely have fewer resources available to allocate for storage of new word information to long-term memory. Providing AV cues during word learning to children with CIs may reduce the cognitive demands of the task imposed by decoding degraded input to support word learning.
Cognitive Load Theory

Cognitive Load Theory (CLT; Sweller, 1988; 2011) stems from the field of educational psychology and describes the strain on working memory as it relates to efficient learning or problem solving. CLT draws on prior work regarding working memory (Baddeley, 1992) and proposes that as strain on working memory processes increases, cognitive load should increase, while learning efficiency decreases. Although CLT is primarily applied to pedagogical methodologies, the underlying assumptions are applicable to this study as the goal was to teach participants novel object-label pairs and measure learning outcomes.

CLT revolves around the balance of extrinsic and intrinsic cognitive load and the demands each place on working memory. Here, intrinsic and extrinsic load each refer to respective elements of the learning material and task. Intrinsic cognitive load refers to the complexity of the material to be learned and is relative to the learner. Thus, to change the intrinsic load, the material needs to be altered to exceed, meet, or fall short of the expertise of the learner, and/or the learner needs to acquire different skills to meet the complexity of the material. Extrinsic (external) cognitive load pertains to the method used to convey the material to be learned. Because extrinsic cognitive load refers to the task, or pedagogical method, it is therefore malleable and can be altered to best meet the learning needs of an individual (Sweller, 2011).

Intrinsic and extrinsic cognitive load demands interact to create a lesser or greater demand on the learner. The cognitive demand can be altered by spatially or temporally separating the presented learning materials or by presenting multiple or few materials at once. Both spatially separating materials and increasing the number of materials presented should increase the cognitive demand on the learner. Per CLT, requiring a learner to engage in this type of split-attention task could hinder learning. A split-attention task is one where the learner needs to attend to and integrate multiple sources of input to understand and learn the material (Ayers & Sweller, 2005). Due to the demands placed on working memory from attending to multiple sources of information, the cognitive load could become too taxing and prohibit effective learning.

In an AV speech perception task, both auditory and visual information is transmitted from the speaker to the listener; however, these cues are neither spatially separated nor temporally desynchronized. So, although AV speech perception is a type of split-attention task, the two streams of information do not compete. Rather, auditory and visual speech are complementary, redundant, and spatially approximate as both signals stem from the same speaker. As such, AV speech information should support a single multimodal perception by decreasing intrinsic and extrinsic cognitive load.

The Intersensory Redundancy Hypothesis

The Intersensory Redundancy Hypothesis (IRH; Bahrick et al., 2004) proposes a framework for understanding the perceptual benefit of redundancy across multiple
sensory modalities early in development. The IRH proposes a “two are stronger than one” perspective where if signals in the environment are temporally synchronized across two senses (e.g., audition and vision), more attention is devoted to that redundant information than if the event were unimodal (e.g., audition only).

In infancy, the redundancy of multimodal information within the environment has been suggested as a tool to guide early learning (e.g., Bahrick & Lickliter, 2000; Bahrick et al., 2010; Jordan et al., 2008). This intersensory redundancy functions to recruit infant attention and therefore facilitate better learning and engagement than unisensory cues. Houston and colleagues (2001) investigated the usefulness of intersensory redundancy for speech sound-video associations in infants with NH and infants with CIs, and propose that attending to temporally-synchronized events, across sensory modalities, may be an indication for later word learning skills. Two stimulus sets across each of two speech sounds (/i/ and /a/) were created and paired with a synchronized video clip. For instance, the word “hop” was paired with a video of a kangaroo hopping up and down and the falling phoneme /i/ was paired with a video of a ball rolling down a ramp. Using the infant intermodal preferential looking paradigm (Hollich et al., 2000) children were familiarized with one of the two pairs of speech sound-video pairs (pairs of either /i/ or /a/ stimuli and their corresponding videos) and to the task. Following stimuli familiarization and task training, participants moved on to testing. Here, both videos were presented side-by-side on the screen, but only one corresponding auditory stimulus was presented. Infant gaze shift to the correct video and amount of time spent looking to each video were recorded.

Children with NH spent a greater amount of time attending to the synchronized video as compared to the distractor video across the age groups (6, 9, 18, and 30 months). Children with CIs displayed variable outcomes and were divided into subgroups for analyses: earlier-implanted and later-implanted. Children who were ‘early implanted,’ meaning they received their CI by age 15 months, demonstrated a gradual shift to the synchronized sound-video pair in accordance with listening experience where amount of time spent looking to the correct video increased in with increased listening experience. Children who were ‘late implanted’ (received their CI by age 24 months) demonstrated a similar gradual shift to increased looking time to the synchronized sound-video pair as listening experienced increased, but this looking time increase was to a lesser degree than what was recorded for the early-implanted participants. The outcomes of this work suggest that infants with CIs are sensitive to stimuli which are temporally-synchronized across sensory modalities (auditory and visual), but the degree to which they attend to temporally-synchronized events may differ with experience.

**Overview of Objectives**

The current study implemented a novel word learning task with both AV and AO speech conditions with school-aged children. The primary objective was to understand the usefulness of combined AV speech information for novel word learning among children with CIs and children with NH. The secondary objective was to quantify looking
time to designated locations on the presentation screen during learning. Outcomes of this work may be used to guide future learning support interventions for children with hearing loss who use CIs.

Research Questions and Hypotheses

Per CLT (Sweller, 1998; Sweller, 2011), both the AV and AO novel word learning tasks should have low cognitive demand. During the AO task, participants were required to listen to an auditory stimulus while a static object was presented on the screen. During the AV task, auditory and visual signals were temporally-synchronized and spatially approximate. So, participants could capitalize on the redundant visual speech information while the static object was also displayed on the screen. When also considering the IRH (Bahrick et al., 2004) and prior word learning work with younger children (Houston et al., 2001), participants should display greater learning during the AV task because it consists of temporally-synchronized, bimodal information. Per the IRH, synchronized bimodal information should recruit greater attention from a listener than unimodal information (i.e., AO). Research questions and hypotheses are outlined below.

Research Question One and Hypothesis

How does access to AV speech information impact novel word learning success for children with CIs, children with NH listening to non-simulated speech, and children with NH listening to CI-simulated speech, when compared to novel word learning in an AO condition? Within the framework of the current study task, participants will have access to speaker’s face during the AV task or only auditory information during the AO task. No other distractors, such as background noise, were presented. Per the IRH and assumptions of CLT, all participants should demonstrate an increase in novel word learning when presented with the speaker’s face (AV) as compared to only listening to the presented information (AO).

Research Question Two and Hypotheses

How do patterns of visual attention during learning relate to individual word learning outcomes for children with CIs and children with NH? Overall, children with CIs should spend more time attending to the speaker’s mouth when compared to other areas of interest as well as their NH peer groups. Children with CIs receive degraded auditory input through their CIs and have demonstrated reliance on speechreading and visual speech information during speech perception tasks in prior works (e.g., Bergeson et al., 2003; Kirk et al., 2007). Children with CIs should also spend more time attending to the speaker’s mouth because it provides intersensory redundancy, which should support novel word learning.
In general, children with NH should spend less time than children with CIs attending to the speaker’s mouth and perhaps more time looking at the speaker’s eyes or object because children with NH have typically-developing peripheral and central auditory systems and should not need to spend most of the trial attending to the speaker’s mouth for supplementary speech input.

**Research Question Three and Hypotheses**

What measured factors (e.g., hearing history, device characteristics, and maternal education level) contribute to novel word learning success across the AV and AO tasks? Standard measures of receptive vocabulary and phonological knowledge should be positively correlated with word learning success. It was also hypothesized that children who were implanted earlier in life would show better word learning outcomes than children who receive their implants later in development.
CHAPTER 2. LITERATURE REVIEW

Speech Is a Multisensory Signal

In everyday situations we commonly have access to both the voice and the face of the person we are listening to, and this is advantageous as speech perception improves with access to AV cues, when compared to AO (e.g., Erber, 1975; Sumby & Pollack, 1954). Prior studies have reported a superadditive effect where AV speech perception performance is greater than the sum of speech perception performance across AO and VO tasks (Tye-Murray et al., 2016). In addition to perceptual accuracy, access to AV speech supports more accurate rates of discrimination and can reduce the subjective effort expended to perceive the speech signal (Fraser et al., 2010).

Auditory and Visual Speech Cues Work Together

Auditory and visual speech cues are statistically related to each other. As the vocal tract changes to produce various auditory outputs, visual oral motor changes occur in accordance. These synchronized changes are temporally aligned and spatially approximate as each signal stems from the same source. The benefit of combined auditory and visual speech information is observed both in quiet and in noisy environments (Helfer & Freyman, 2005; Schwartz et al., 2004; Sumby & Pollack, 1954). In background noise specifically, visual speech information can function to help resolve an ambiguous auditory signal, or competing auditory signals, by providing information that can be difficult to resolve with auditory cues alone. For instance, place cues are visual indicators of where a sound was produced in the vocal tract. The higher the production in the vocal tract (e.g., lips), the easier the production is to see. Visible place cues can function to distinguish similar phonemes to improve perceptual accuracy. As an example, the phoneme /b/ (as in ball) is produced at the lips whereas the phoneme /d/ (as in doll) is produced further back in the vocal tract, at the alveolar ridge. Watching the speaker’s oral motor articulations during the production of the word ball supports the listener’s perception of the word ball and helps prevent confusion with the word doll.

In the late 1990s, Yehia and colleagues reported a predictive relationship between changes in the vocal tract, corresponding changes in facial movement, and auditory output (Yehia et al., 1998). More recently, Chandrasekaran and colleagues (2009) assessed statistical correspondences across AV speech productions from multiple databases which included stimuli from American English, British English, and French speakers. The speech productions across these databases were analyzed for acoustic output as well as lip-to-lip distance during productions. Consistent with previous reports (Yehia et al., 1998), the authors described predictable relationships between the visual articulators and acoustic output across English and French. These statistical correspondences included a temporal relationship between the opening of the mouth and the acoustic envelope of the signal as well as a relationship between the timing of the speaker’s mouth movements and the onset of the acoustic signal.
Listeners are able to capitalize on these statistical, predictable relationships between auditory and visual speech cues to optimize speech perception. One notable cue that may be consistent across auditory and visual speech is neighborhood density. Neighborhood density, in the auditory modality, refers to the number of words that differ from a target word by one phonemic addition, substitution, or deletion. Words in a dense auditory ‘neighborhood’ have many other words within a language that minimally differ from them and conversely, words in sparse ‘neighborhoods’ have few other words in a language that minimally differ from them. Neighborhood density can influence word recognition speed and accuracy where words in a dense neighborhood are more slowly recalled and recalled incorrectly more often, as compared to words in a sparse neighborhood, presumably due to a larger number of competitor words (Luce & Pisoni, 1998). Tye-Murray and colleagues (2007) have proposed that visual speech cues may be associated with one another in a similar manner as auditory neighborhood density cues. Where auditory neighborhood density refers to a group of words that minimally differ by one phoneme, visual neighborhood density refers to a group of words that minimally differ by the way they look when produced.

In an AO scenario, recognizing a word in a dense auditory neighborhood should be slower and less accurate than if that word was in a sparse neighborhood (Vitevitch & Luce, 1999). However, in an AV scenario, the listener has access to visual cues in addition to auditory, which may interact to influence speech perception. Thus, if the same high density word is presented in an AV scenario, the visual neighborhood density of that word can constrain the number of competitors and alter recognition speed and accuracy. So if the word is in a dense auditory neighborhood, but sparse visual neighborhood, the combined AV cue should afford faster and more accurate recognition than a word in a dense auditory and dense visual neighborhood.

To investigate the influence of auditory and visual neighborhood density on speech recognition, Tye-Murray and colleagues (2007) presented words that varied by density characteristics to 131 NH adults across AO, VO, and AV conditions. Participants demonstrated poorer recognition of words in a dense auditory neighborhood in the AO condition. As hypothesized, participants also demonstrated poorer word recognition of words in dense visual neighborhoods in the VO condition, suggesting visual neighborhood density can also influence speech perception. In the AV condition, words with minimal overlap across auditory and visual density neighborhoods were better recognized than words with greater overlap. Thus, words with auditory neighbors that are visually dissimilar were better recognized. This work supports the notion that auditory and visual speech cues work together to support speech perception and that listeners are sensitive to the relationship between auditory and visual speech cues.

One of the earliest studies to investigate the benefit of access to visual speech cues demonstrated marked improvement in speech perception in noise when the listener could see the speaker while listening (AV) compared to listening without visual speech cues (AO). More than 100 NH adults were presented spondee words, in quiet and in noise, at varying signal-to-noise ratios (SNRs). Word stimuli were presented in AO and AV formats. Speech perception worsened as the SNR decreased, or as the background
noise grew louder than the spondee word, across conditions. However, when participants had access to visual speech cues (AV), speech perception performance improved significantly from the AO condition. Additionally, the speech perception enhancement experienced during the AV condition was greatest for the worse listening conditions, or for the lowest SNR conditions (Sumby & Pollack, 1954). This study demonstrated the profound impact visual speech information can have on perception and demonstrates the importance of visual speech cues, particularly in high noise/low signal listening situations.

**McGurk Task**

Subsequent early work elaborated on the impact of visual speech, beyond the perceptual benefits, to show that visual speech information can not only enhance, but actually alter our auditory perceptions. In a seminal study conducted by McGurk and MacDonald (1976), changes in the congruency of combined auditory and visual speech information can alter the auditory percept experienced by the listener. In the now termed ‘McGurk effect’ listeners are tasked with perceiving consonant-vowel (CV) syllables (e.g., /ba/, /ga/, /ka/, /pa/), presented as pairs (e.g., /ba/-/ba/ or /ba/-/ga/) in AO and AV conditions. During AO trials, participants listen to the CV pairs and repeat what they hear with no visual information. During AV trials, listeners are asked to watch and listen to a speaker produce the CV pairs and repeat what they hear, but during AV trials, the auditory and visual cues may be incongruent (e.g., auditory /ba/ - visual /ga/). Across these incongruent cue trials, many listeners report neither the perception of the acoustic cue (/ba/), nor the visual cue (/ga/), but rather a “middle” or fused percept, /da/. This fused percept is understood as integration of input from the auditory and visual systems to achieve accurate speech perception and suggests that visual speech information is a significant contributor to speech understanding.

Versions of the McGurk task (McGurk & MacDonald, 1976) have been expanded to listeners with CIs (e.g., Rouger et al., 2008; Schorr et al., 2005). Schorr and colleagues presented the McGurk task to children with NH and children with CIs (M = 5.85 years). Using /pa/ and /ka/ syllables, the authors created congruent (e.g., auditory /pa/ - visual /pa/) and incongruent (e.g., auditory /pa/ - visual /ka/) stimuli, which were randomly presented across 70 trials. Children with CIs demonstrated accurate perception of each syllable in congruent trials; however, following incongruent trials, participants demonstrated a visual dominance effect where the percept reported was skewed to the visual cue presented, not the auditory cue or the fused percept, often reported by people with NH (Schorr et al., 2005).

**Multisensory Speech Perception Across Development**

Auditory-visual speech cues play a significant role in our perception of speech and do so from very early in life, even prior to the acquisition of spoken language. Early works reported an infant preference for temporally-synchronized AV speech signals. Dodd (1979) presented nursery rhymes that were either temporally-synchronized or offset
from the speaker’s oral articulations (misaligned by 400ms) to ten- to sixteen-week-old infants. Looking time to each of the stimuli was recorded. Infants were more attentive to during synchronized trials than to offset trials demonstrating a preference for congruent AV information.

In a seminal study, Kuhl and Meltzoff examined the looking preferences of 18-20-week-old infants to matched and mismatched vowel stimuli. Infants were situated in front of a screen which displayed two faces, side-by-side, while listening to a female speaker produce two vowel stimuli, /a/ and /i/. Vowels were presented with the articulating faces across matched (i.e., articulating /a/ with /a/ vowel) and mismatched (i.e., articulating /a/ with /i/ vowel) trials. Infants spent a greater proportion of time looking to the face during matched trials as compared to mismatched (Kuhl & Meltzoff, 1982). These results were replicated in a subsequent study using the vowels /i/ and /u/ (Green et al., 1988).

Patterson and Werker extended the work presented by Kuhl and Meltzoff (1982) to include female and male faces and voices in an older cohort of infants. Four and a half month-olds were presented with female and male faces situated side-by-side on a screen in front of them and the vowels /a/ and /i/ were used as stimuli across matched and mismatched trials. The outcomes of this work were consistent with prior studies (Green et al., 1988; Kuhl and Meltzoff, 1982; 1984) demonstrating an infant preference in looking to matched stimuli (i.e., the speaker was articulating the vowel presented) as compared to mismatched (Patterson & Werker, 1999). This finding has also been observed in infants as young as two months of age (Patterson & Werker, 2003).

Recent work has continued beyond perception to investigate visual speech information contained in lexical knowledge. Specifically, whether visual speech information is associated with known words in a child’s lexicon. Havy and Zesiger (2017) investigated the visual component of lexical representations in 40 NH children ($M = 30$ months of age). The authors presented a novel word learning task which consisted of pre-familiarization, learning, and test phases. During the pre-familiarization phase, participants were familiarized with each of the object-label pairs in all possible pairings. During the learning phase, participants were broken into two groups, auditory and visual, which represented each learning phase stimulus presentation modality. Children in the auditory learning group saw the novel objects on a black screen, one at a time, while the label was presented via loudspeaker. Children in the auditory group did not have visual access to the speaker. Children in the visual learning group saw the speaker labeling each of the novel objects, but did not have access to the auditory speech information. During the test phase, each of the newly learned objects were presented on the screen side-by-side and participants were asked to select one of the objects based on a prompt. Testing occurred across same and opposite modality trials. During same modality test trials, the test prompt was delivered in the same modality as was presented in the learning phase. During opposite modality trials, participants were prompted with the opposite modality as they received during the learning phase.

Participants who received an AO presentation during the learning phase demonstrated learning of the novel object-label pairs during both same and opposite
modality test trials, as indicated by greater looking time to the target, compared to the distractor. Participants who received a VO presentation during the learning phase demonstrated learning of the novel object-label pairs during the same modality test trials, but not during opposite modality test trials. The results of this study support the idea that young NH children are able to integrate visual speech information with their understanding of new words even without visual access to the speaker.

In another recent study of AV speech perception in preschool-aged (30-48 months old) NH children, Grieco-Calub (2015) sought to understand the influence of central factors on speech processing. Target object presentations were given in quiet and in noise, across AO and AV speech conditions. Reaction time to looking at a target object and amount of time spent looking at the target were measured. Participants who were slower to look at the correct object in the AO condition demonstrated benefit when provided visual speech information (AV), but participants who were quick to orient to the correct object during the AO condition did not demonstrate a significant benefit when provided additional visual cues (AV). Outcomes of this work suggest that additional visual speech cues may help support speech perception, particularly for listeners who show difficulty with speech perception in an AO context.

U-shaped Developmental Trajectory

Although infants and young children seem to prefer visual speech cues (e.g., Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; 2003) and integrate visual speech information when learning new words (Havy et al., 2017; Havy & Zesiger, 2017), the influence of visual speech on perception has been proposed to change across development. Jerger and colleagues have described a ‘U-shaped’ trajectory of development for AV integration across the lifespan (Jerger et al., 2009). Across this trajectory, infants and young children demonstrate AV integration during speech perception tasks, then around five-to-nine years of age, demonstrate a lesser reliance on visual speech information. After age ten and into adulthood, a strong influence of visual speech information seems to emerge again (Jerger et al., 2009). This proposed U-shaped developmental trajectory results from outcomes of studies of AV speech perception tasks in this preschool to early elementary school age group (e.g., Hockley & Polka, 1994; Lalonde & Holt, 2015; Massaro, 1984).

Lalonde and Holt (2015) investigated AV speech perception among three- and four-year old NH children as well as a group of NH adults. Preschool-aged children completed a battery of experimental measures including recognition, matching, and distinguishing between stimuli. During recognition, participants were asked listen to words from the Lexical Neighborhood Test (Kirk et al., 1995) in background noise across AV and AO conditions then repeat what they heard. All groups of participants demonstrated higher accuracy during AV trials as compared to AO. In the matching task, participants heard nonsense syllables while watching two articulating faces, presented side-by-side on a screen. Participants were asked to indicate which face matched the auditory stimulus. Adults demonstrated ceiling-level performance whereas children demonstrated highly variable performance with the younger participants (three-year-olds).
close to chance-level. During the distinguishing task, participants were presented pairs of stimuli and were asked to indicate if the two stimuli within ear pair were the same or different. Three- and four-year-olds completed the distinguishing task in AV and AO conditions, and minor modifications were made during the distinguishing task across age groups to ensure each participant could complete the task. Participants demonstrated similar rates of discrimination during the AO condition. In the AV condition, children demonstrated better discrimination when the speech tokens were more visually salient, or more visible during articulation. Outcomes across tasks in this study suggest young children, between three and four years of age, can integrate visual speech information during perceptual tasks.

Lalonde and Holt found comparable results in a study of older NH children (Lalonde & Holt, 2016). Six- to eight-year-old children completed three speech recognition tasks: detection, discrimination, and recognition. The detection task consisted of AO and AV conditions. During the AO detection task, either noise was presented, or a word stimulus was presented within the noise. Participants were instructed to indicate whether or not a word was present in the noise. The AV detection task consisted of match and mismatch trials. During match trials, if a word was presented in the background noise, a corresponding video was present on the screen and if there was no word embedded in the background noise, a still image was presented. Conversely during mismatch trials, if a word was present in the background noise, a still image was presented, and if no word was present, a video was presented. The discrimination task also consisted of AO and AV trial types. During AO trials, participants listened to pairs of words from the Lexical Neighborhood Test (Kirk et al., 1995) in noise and indicated if the two words were the same or different. AV discrimination trials were either matched or mismatched. During the recognition task, participants were presented a word from the Lexical Neighborhood Test in noise followed by a brief silent interval, then a probe stimulus. Participants were asked to indicate whether or not the presentation in noise matched the probe and presentations were made in AO and AV conditions across matched and mismatched trials. Children showed highest performance during AV as compared to AO conditions; however adults demonstrated a greater sensitivity to the addition of visual cues when compared to the child participants, across tasks.

Recently, Jerger and colleagues (2018) investigated AV speech recognition performance across a wide age range of participants (4-14 years of age). They presented a monosyllabic nonsense word ‘buh’ to 115 children across AO, VO, and AV conditions. VO and AV conditions contained two different trial types, one where the talker was speaking dynamically and another where the talker was static. Participants were instructed to indicate each time they saw and/or heard the person articulating the nonsense word “buh.” Older participants showed much faster response times than younger participants across AO, VO, and AV trials. Only the youngest group of participants (four-to-five years of age) demonstrated faster reactions to the dynamic speaker in the AV condition. Older participants did not show a response difference for the dynamic speaker across AO and AV trials. Across age groups, reaction times were faster across AO and AV trials as compared to VO trials. Outcomes from this study highlight differences in AV speech perception across development.
Auditory-Visual Integration in Children with Cochlear Implants

These descriptions of normal multisensory development assume that information is transmitted through healthy and typical auditory and visual systems, but how does this change when one of the primary senses that contributes to multisensory perception of speech is developing differently than what is observed in typical listeners?

Cross-Modal Reorganization

Children with CIs experience a period of auditory deprivation prior to receiving their device(s). Auditory deprivation can lead to neuronal changes at the level of the cortex where neural connections responsible for transmitting information to the deprived auditory modality are recruited to other areas of the brain, often the visual cortex (Campbell and Sharma, 2016; see Anderson et al., 2007 for review). This cross-modal reorganization (CMR) within the cortex is commonly deemed responsible for the visual speech bias reported across listeners with CIs during speech perception tasks (Desai et al., 2008; Stropahl & Debener, 2017). Much of what we know about the perceptual consequences of CMR for listeners with CIs stems from research which compares evidence of CMR to speech perception abilities. Across these studies, evidence of CMR is associated with poorer speech perception abilities in adults (e.g., Doucet et al., 2006; Lomber et al., 2010) and children who use CIs (Bergeson et al., 2010; Campbell & Sharma, 2016).

Unisensory and Multisensory Perspectives

Due to prolonged auditory deprivation, children with CIs are assumed to have experienced some degree of CMR. CMR may have downstream effects on speech perception for children with CIs (Sharma et al., 2015) and differing perspectives exist as to the best way to capitalize on input via the CI to improve speech and language outcomes. Previous reports have suggested that since auditory neuronal connections are likely recruited to the visual cortex, providing visual information to listeners with CIs would prohibit the development of the auditory pathway following implantation (Champoux et al., 2009). This assumption formed a unisensory perspective of perception which assumes that removing the visual cue (i.e., the speaker’s mouth) will force the listener to attend to the auditory cues from the CI. In doing so, re-establishment of the auditory pathway could occur. Recently, views have started to shift away from a unisensory perspective, to a multisensory perspective, which posits that providing both auditory and visual speech cues best supports speech and language development. This perspective was constructed from demonstrated benefit of combined auditory and visual speech cues across various speech perception tasks (Holt et al., 2011; Kirk et al., 2007; also see McDaniel & Camarata, 2017 for review).
Spectral Degradation

Children with CIs may receive significant benefit from visual speech cues because a CI does afford a rudimentary sense of hearing; however, the transmitted signal is much different than what a NH person receives through natural acoustic hearing. CIs are the most successful neuroprostheses to date and restore auditory perception in listeners with severe-to-profound hearing loss, affording access to spoken language. Externally, the CI is comprised of a transmitter coil, microphone(s), and speech processor. Internally, the CI consists of a receiver and magnet, placed sub-dermally in the mastoid bone, and electrode contacts along an array, inserted into the cochlea. The external speech processor “picks up” sounds from the listener’s environment and converts the acoustic information to a digital signal. The digital information is transmitted to the internal implant, or receiver, where it is transcribed to electrical impulses. The impulses are transmitted to the electrode contacts along the array where the contacts stimulate various regions along the basilar membrane. Once stimulated, the brain is able to interpret the electrical impulses as sound.

CI signal transmission is a multi-step process. Broadly, the acoustic signal is divided into frequency bands, then the envelope of the signal is extracted. Each band is then transmitted to its respective frequency place along the electrode array. Oftentimes, when stimulation is delivered, the electric current spreads to other areas of biological tissue, unintentionally. The current spread can then stimulate adjacent neurons and distort speech perception. Although current CI technology allows for more than 20 intracochlear electrodes, most listeners with CIs only receive 6-10 spectral channels of information (Fishman et al., 1997; Garnham et al., 2002). In addition to the spreading of electric current, the CI channel constraint problem limits access to spectral information. This spectral degradation can impair accurate perception of vowels and distort perception of many consonant place and manner characteristics (Nie et al., 2006; Shannon et al., 1995). Visual speech cues can function to help listeners with CIs overcome the degraded input signal to achieve accurate speech perception.

Auditory-Visual Speech Perception and Higher Order Cognitive Skills

The majority of published work that contributes to the understanding of AV speech perception among children with CIs describes word or sentence recognition across AO, VO, and AV tasks. Kirk et al. (2007) assessed AV speech perception in 15 children with CIs. The purpose of this study was not only to assess AV speech perception, but also the influence of lexical characteristics on AV speech perception. Fifteen children with CIs ($M = 5.89$ years) listened to words from the Lexical Neighborhood Test (Kirk et al., 1995) as well as sentences from the AV-Lexical Neighborhood Sentence Test (Holt et al., 2005) in AO, VO, and AV conditions. The lexical characteristics of the words across these measures varied such that lexically easy words had high probability-low density characteristics and lexically difficult words had low probability-high density characteristic. Performance during the AO condition across both word and sentence recognition tasks was widely variable, so the authors divided participants into two groups based on outcomes, ‘good’ and ‘poor’ performers. Participants were separated into their
respective groups according to performance on the AV-Lexical Neighborhood Sentence Test.

In the AO conditions, participants in the ‘good’ performers subgroup demonstrated better performance (higher percent correct) when perceiving sentences as compared to words whereas ‘poor’ performers showed better performance when perceiving words as compared to sentences; however, performance was still lower than that of the ‘high’ performers. Both groups of participants were better able to perceive lexically easy words as compared to lexically difficult words. Across speech perception tasks, all participants demonstrated highest performance overall in the AV condition as compared to AO or VO.

Lachs and colleagues (2001) presented sentences in AO and AV contexts to 27 children with CIs to assess AV benefit, or the perceptual gain achieved from adding visual cues to auditory information as compared to auditory information alone. The authors assessed perception of sentences from the Common Phrases Test (Osberger et al., 1991) across AO, VO, and AV conditions as well as the Lexical Neighborhood Test, Multisyllabic Lexical Neighborhood Test (Kirk et al., 1999), and the Phonetically Balanced Kindergarten word lists (Haskin Labs). Receptive vocabulary and speech production were also assessed. AV benefit, or perceptual gain was measured by comparing the AV score to the A score to attain the benefit of adding visual speech cues. As reported in Kirk et al., 2007, performance in this study was widely varied so the authors sub-grouped participants by ‘good’ and ‘poor’ performers. The sub-grouping was based on the median score for each of the AO measures. These subgroups demonstrated significantly different performance outcomes on the lexically easy words in the Lexical Neighborhood Test, on both word types in the Multisyllabic Lexical Neighborhood Test. Additionally, the addition of visual speech cues did not allow a poorer performer to become an excellent performer as children who were ‘good’ performers in the AO conditions gained the most benefit from the addition of visual speech cues.

AV integration may not only provide information about speech perception for children with CIs, but may inform later language development as well. Bergeson and colleagues (2005) used a battery of assessments, given at multiple time points, to characterize speech and language skills in 80 children with CIs. The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) was used as the measure of receptive vocabulary knowledge and the Common Phrases test, given in AO, VO, and AV contexts, was used as a measure of AV speech perception. In accordance with implant use and listening experience, participants demonstrated an increase benefit of combined AV speech cues, but visual gain decreased over time for children who used oral communication. Importantly, the authors found that pre-implant outcomes on the Common Phrases test, in VO and AV conditions, were correlated with vocabulary knowledge years after implantation. This outcome would suggest that AV speech perception skills can be indicative of later vocabulary development.
Word Learning Task

Early AV integration ability may be indicative of later language skills among children with CIs, including word learning (Bergeson et al., 2005; Houston et al., 2001); however, very little work has been done to investigate the influence of AV integration on novel word learning. Most published work has assessed speech perception outcomes in AO and AV word and sentence recognition tasks (e.g., Kirk et al., 2007; Lachs et al., 2001). The task demands of a word or sentence recognition task differ when compared to a novel word learning task, so the benefit of AV speech cues across these task types may differ in accordance. Speech perception requires attending to a signal and retaining this information in short-term, and possibly long-term, memory until recall, whereas word learning requires creation of new word knowledge, and occurs over a series of stages. The learner needs to be able to identify a novel word, rapidly associate the new label with its referent, then refine this new word knowledge over time, in accordance with additional exposure, and retain this information in long-term memory. Therefore, the word learning process requires large attentional and working memory resources.

Often discussed as one of the first steps of word learning is fast mapping (Carey & Bartlett, 1978), which describes the process of rapidly associating a word with its referent after very brief exposure, and has been proposed as the initial step in acquiring a new word (Ellis Weismer & Evans, 2002; also see Kucker et al., 2015 for review). Knowledge of a fast mapped word may contain fragmented pieces of information regarding the word’s phonology, meaning, category, and syntactic structure. Additionally, these representations may also be as detailed as to contain information regarding the environment or context around which the word and its referent were first encountered (Dollaghan, 1987). The representation created during fast mapping is not solidified. In fact, the representation is often lacking in information and may not be completely accurate; however, the fast mapped representation is enough to trigger recall when the word is next encountered so the listener can refine and solidify an accurate representation, over time, in accordance with additional experience. Some early hypotheses about fast mapping proposed listeners can use the process of elimination to infer meaning. For example, given the statement “Pick the yellow flower, not the magenta one.” If the listener knows the meaning of the word yellow, but not magenta, then she can infer that yellow and flower are not part of the meaning of magenta and she can additionally infer, if magenta is a color word, it does not denote the color yellow (Markman & Hutchinson, 1984). Additionally, fast mapping has been discussed as a tool that young children can use to quickly acquire new vocabulary words (Carey, 1978).

Following identification of a word as novel, the phonological, lexical, and semantic characteristics need to be integrated with the listener’s established word knowledge. These associations are formed over time as the listener gains more experience with the word and in varied contexts (Kucker et al., 2015). Retaining newly learned information in memory is essential for building a vocabulary. Upon initial fast mapping of the label and referent, the learner needs to retain this new information in memory until it can be retrieved when the word is next encountered in order to build and solidify the new word knowledge.
Vocabulary Skills of Children with Cochlear Implants

Vocabulary acquisition may become problematic if a breakdown occurs across the stages of word learning and children with CIs show some difficulties with acquiring new words, specifically with retaining newly learned information. Walker and McGregor (2013) examined the word learning process in 24 children with CIs ($M = 4.86$ years of age). The authors taught novel object labels, for eight novel objects, to each participant across two sets. After training, the authors probed participant knowledge with a series of questions to assess if the child could remember the name of the novel objects and to assess if the participants could extend the labels to other objects within a category (same object, but differed in either size or color). Participants were probed at a second visit to assess word retention and all participants demonstrated learning or were able to indicate the labels for the novel objects presented. Children with CIs and their NH vocabulary-matched peers scored similarly on the extension measure, but the age-matched NH group showed ceiling performance. Similarly, the CI and vocabulary-matched NH participants had poorer retention of learned words than the age-matched NH group. Thus, children with CIs seem to perform more poorly than their same-age NH peers on assessments that address multiple stages of the word learning process.

Disparate and widely variable word learning outcomes have been reported across various studies of children with CIs when compared to children with NH (e.g., Geers, 2002; Geers et al., 2003; Geers et al., 2009; Houston & Miyamoto, 2010; Sarant et al., 2001; Walker & McGregor, 2013). Oftentimes these differences in development are attributed to the auditory deprivation experienced by children with severe-to-profound hearing loss prior to receiving their CI(s). Once children receive auditory input via the CI, they would need to acquire new words at a rapid rate to ‘catch up’ to their NH peers. Hayes and colleagues (2009) investigated receptive vocabulary knowledge and growth rate in 65 children with CIs. Receptive vocabulary was measured via the PPVT (Dunn & Dunn, 1997) once per year, over three consecutive years. Over time, receptive vocabulary scores improved and rate of improvement varied by age-of-implantation where children who were implanted earlier showed faster rates of improvement than children implanted later in life. More recent studies of novel word learning have also indicated that age-of-implantation is a significant contributor to word learning success (Geers & Nicholas, 2013; Geers et al., 2009; Houston et al., 2012).

Although earlier age-of-implantation may support novel word learning, many children who receive CIs early in life still struggle to attain age-appropriate language and vocabulary skills. Svirsky and colleagues (2004) assessed speech perception at multiple time points, around initial activation of the device, and again two to eight times following device activation. Participants were sub-grouped by age-of-implantation and speech and language skills were assessed via the Reynell Developmental Language Scales (Edwards et al., 1997) and the MacArthur-Bates Communicative Development Inventories (Fenson et al., 1993), which is a parent-report indicator of receptive vocabulary knowledge. Children in the youngest age-of-implantation group, those implanted by 24 months of age, did not attain speech, language, and vocabulary outcomes equivalent to their NH peers at the same trajectory rate. Children in this age group required an additional year to
reach equivalent performance levels of the NH participants. Children implanted after 24 months of age demonstrated poorer performance and did not reach NH performance within the time frame of this study.

Davidson and colleagues (2014) moved beyond age-of-implantation to assess the influence of audibility on novel word learning success. Audibility was measured through aided thresholds attained in the sound-field. A low-frequency pure tone average (PTA) and high frequency PTA were calculated. Measures of speech perception, receptive vocabulary knowledge, and novel word learning were also conducted. Participants with better audibility (lower PTAs) demonstrated greater receptive vocabulary knowledge and better novel word learning than participants with poorer audibility (higher PTAs), although the difference was not statistically significant. Overall, children with CIs performed poorer on the vocabulary and learning measures than children with NH.

Recently, Lund (2015) conducted a meta-analysis to assess multiple variables that may contribute to word learning outcomes in children with CIs to afford a better understanding of vocabulary development within this group. The analysis revealed that children with CIs are unable to ‘catch up’ to the vocabulary skills of their age-matched NH peers, which is likely attributed to the delay in language exposure children with CIs experience prior to implantation.

Although some children with CIs are able to attain age-appropriate vocabularies, all children with CIs learn new words at a slower rate than their NH peers, and some exhibit persistent language delays and never attain appropriate vocabulary skills (Geers et al., 2016; Lund, 2015). Numerous factors have been investigated as potential sources of individual variation for language outcomes among listeners with CIs, including amount of time spent listening with CIs (Nicholas & Geers, 2007), participation in an auditory-aural (re)habilitation program (Hayes et al., 2009), number of intracochlear electrodes (Blamey et al., 1992), etiology of the hearing loss (Gantz et al., 1994), amount of residual hearing prior to implantation (Niparko et al., 2010), implant technology (Geers et al., 2003), use of oral communication (Geers, 2002; Sarant et al., 2001), and duration of deafness (Geers, 2003). Higher rates of word learning among children with CIs are most often attributed to earlier age of implantation and greater experience listening with the device (e.g., Hayes et al., 2009; Houston et al., 2012), but much of the variance in CI word learning outcomes is unaccounted for.

**Auditory-Visual Word Learning Among Children with CIs**

Few studies have explicitly examined the impact of access to visual speech cues on novel word learning in children with CIs. Lund and Schuele (2017) assessed the influence of AV cue synchrony on novel word learning performance in nine children with CIs and age-matched children with NH. Here, AV cue synchrony was in reference to labeling of the object and indicating the labeled object, not auditory and visual speech cues. Thus, during synchronous trials, the object was manipulated and labeled by the researcher at the same time. During asynchronous trials, the experimenter labeled the
object, then after a brief delay, manipulated the object. Participants were instructed to identify a target object following a delay between presentation and test. Children with NH demonstrated overall better performance than children with CIs with best performance for synchronous presentations as compared to asynchronous. Children with CIs demonstrated comparable performance across synchronous and asynchronous presentations.

In a recent study, McDaniel and colleagues (2018) taught novel object-label pairs to four children with CIs (4-5 years of age), using AV or AO instruction, to assess the influence of visual cues on novel word learning success. The authors worked with each participant multiple times per week to teach novel object-label pairs and probe learning of the object-label pairs using AV or AO instruction. During the AV condition, participants had access to the speaker’s mouth. During the AO condition, speechreading cues were not available. Participants demonstrated learning of the object-label pairs across conditions (AV and AO), but learning was not significantly better in either condition. Due to the limited sample size it is unclear if this pattern would be observed in a larger, more varied group of children with CIs.

The Current Study

Although speech is a multimodal signal and children spend most of their day speaking with teachers and peers, the role of AV information during vocabulary acquisition in the school-aged CI population is not well understood. Moreover, while pediatric CI listeners appear to benefit from AV information for speech perception tasks (e.g., Bergeson et al., 2003; Lachs et al., 2001), much less is known about the role of AV information in more complex tasks, such as novel word learning. Not only is an understanding of the efficacy of AV speech cues important to improve our understanding of vocabulary acquisition, but it may also be critical for supporting later language skills (Bergeson et al., 2005).

Developing an expansive vocabulary relies on successful word learning (Houston et al., 2001). To support lexical development, multiple processes must work together, including auditory mechanisms responsible for perception, as well as language knowledge, with sufficient memory and cognitive capacity to attend to and retain word knowledge (Gathercole et al., 1997). Due to the inter-reliance of these processes, deficiencies in any one could cause difficulty building a robust lexicon. As such, children with CIs, must overcome huge obstacles to achieve appropriate vocabulary development. While facing these challenges, children with CIs still must learn new words every day in multi-modal environments. Therefore, focusing on how children with CIs use visual speech information during the vocabulary acquisition process will help foster an understanding about how these children build their vocabularies in the real world.
CHAPTER 3. METHODOLOGY

Overview

The primary goal of this study was to examine novel word learning differences across AV and AO presentations in children with CIs and children with NH, listening to non-simulated (NH-NoSim) or CI-simulated (NH-CISim) speech. The secondary goal of this study was to characterize individual looking patterns as they relate to individual word learning outcomes for children with CIs. The novel word leaning tasks (AV and AO) consisted of three phases, Familiarization, Learning, and Test. During Familiarization, eight familiar object-label pairs were presented as a practice Test task. The Familiarization phase was used to ensure participants did not make errors during testing due to misunderstanding the task. Following Familiarization, the Learning phase was presented. The Learning phase was adapted from Storkel (2001) and was set-up like an electronic storybook that was presented across three blocks. Within each block, eight nonword-object pairs were presented, one at a time. As the story progressed from Block one to Block three, the number of presentations of each nonword increased. By the end of Block three, participants received 10 cumulative exposures to each nonword-object pair. Learning of each pair was assessed after each Learning block, during a Test phase, with a four-alternative forced-choice (4AFC) referent identification task. See Figure 3-1 for a task overview.

Participants

Two novel word learning tasks (AV and AO) were presented to 13, six- to nine-year-old children with CIs and 32, six- to nine-year-old children with NH. One child with a CI was excluded from analyses due to refusal to comply with the protocol and 24 NH participants were chosen for the analyses based on best chronological age and sex match to each CI participant. Thus, 12 children with CIs and 24 children with NH (12 NH-NoSim and 12 NH-CISim) were included in the final analyses. Each group was comprised of seven females and five males. The ages of participants in the CI group ranged from 6.1 years to 9.8 years ($M = 7.86$ years) and the ages of the participants in the NH groups ranged from 6.0 years to 9.9 years months ($M = 7.86$ years).

Participant Recruitment

Participants were recruited from multiple locations across eastern and middle Tennessee through social media, word of mouth, posted and mailed flyers, phone calls, and e-mail announcements.
<table>
<thead>
<tr>
<th>Phase</th>
<th>Example Visual Stimuli</th>
<th>Example Auditory Stimuli (Excerpts)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarization*</td>
<td><img src="image1" alt="Auditory-Visual" /> <img src="image2" alt="Auditory-Only" /></td>
<td>“Where’s the apple?”</td>
<td>Participants were asked to select each of the eight familiar object-label pairs, one at a time.</td>
</tr>
<tr>
<td>Learning Block 1</td>
<td><img src="image3" alt="Auditory-Visual" /> <img src="image4" alt="Auditory-Only" /></td>
<td>“I’ll get my <em>poin</em>” “My favorite is from the <em>dav</em>”</td>
<td>Each of the 16 novel object-label pairs were described one at a time. Eight pairs were presented in each task (AV and AO). Each label was presented one time.</td>
</tr>
<tr>
<td>Test Block 1*</td>
<td><img src="image5" alt="Auditory-Visual" /> <img src="image6" alt="Auditory-Only" /></td>
<td>“Where’s the <em>poin</em>?” “Where’s the <em>dav</em>?”</td>
<td>Participants were asked to select each of the eight familiar novel object-label pairs, one at a time.</td>
</tr>
<tr>
<td>Learning and Test* Block 2</td>
<td><img src="image7" alt="Auditory-Visual" /> <img src="image8" alt="Auditory-Only" /></td>
<td></td>
<td>Following Test Block 1, Learning and Test Blocks 2 and 3 were presented. During Block 2, each object label was presented four times. During Block 3, each object label was presented 5 times.</td>
</tr>
<tr>
<td>Learning and Test* Block 3</td>
<td><img src="image9" alt="Auditory-Visual" /> <img src="image10" alt="Auditory-Only" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Feedback was provided following each response.

**Figure 3-1. Overview of each phase across AV and AO presentations**

Children with Cochlear Implants

Participants with CIs were recruited from two clinics at the University of Tennessee-Knoxville, the audiology clinic and the Child Hearing Services clinic. Additional participants were recruited from Tennessee School for the Deaf and the greater eastern and middle Tennessee area. Table 3-1 displays participant hearing history information including age of initial amplification, age of cochlear implantation, and etiology of hearing loss for each CI participant.

Children with Normal Hearing

Participants with NH were recruited through flyers posted on social media, word of mouth, and paper flyers dispersed around the University of Tennessee-Knoxville campus. In addition to recruitment in the greater Knoxville community, private and public schools in Knox County agreed to support this research project through e-mailed announcements sent to parents and guardians of children who attend those schools and met age criteria. All NH participants completed a hearing screening at 20dB HL across 500, 1000, 2000, and 4000 Hz, bilaterally.

Inclusion and Exclusion Criteria

Participants were recruited based on the inclusion and exclusion criteria outlined below. Information concerning inclusion and exclusion criteria were collected from patient medical charts (CI participants only) and confirmed through parental report.

Inclusion criteria for children with cochlear implants

- At least six months of experience listening with the device(s)
- Normal-to-corrected vision
- 6:0 – 9:11 (years:months)
- Unilaterally or bilaterally implanted with CIs

Inclusion criteria for children with normal hearing

- Normal bilateral hearing thresholds, defined as hearing thresholds at or below (better) 20dB HL at 500, 1000, 2000, and 4000 Hz
- 6:0 – 9:11 (years:months)
- Normal-to-corrected vision

Exclusion criteria for all children

- English is not the primary language spoken in the home
- Developmental or cognitive delay as indicated by a review of the medical chart and/or confirmed via parental report
- Diagnosis of epilepsy and/or history of seizures
Table 3-1. Demographic and hearing history information for CI participants

<table>
<thead>
<tr>
<th>Child</th>
<th>Sex</th>
<th>Age (years. months)</th>
<th>Etiology of HL</th>
<th>Age of HL Diagnosis</th>
<th>Age of HA Fitting</th>
<th>Age of CI (months)</th>
<th>Other Ear Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>F</td>
<td>7.4</td>
<td>Connexin 26</td>
<td>birth</td>
<td>9 months</td>
<td>12</td>
<td>bilateral</td>
</tr>
<tr>
<td>AB</td>
<td>F</td>
<td>7.4</td>
<td>Connexin 26</td>
<td>birth</td>
<td>9 months</td>
<td>12</td>
<td>bilateral</td>
</tr>
<tr>
<td>AC</td>
<td>F</td>
<td>8.9</td>
<td>Unknown</td>
<td>6 months</td>
<td>18 months</td>
<td>24</td>
<td>none</td>
</tr>
<tr>
<td>AD</td>
<td>M</td>
<td>8.6</td>
<td>ANSD</td>
<td>6 months</td>
<td>18 months</td>
<td>84</td>
<td>HA</td>
</tr>
<tr>
<td>AE</td>
<td>M</td>
<td>6.1</td>
<td>Unknown</td>
<td>12 months</td>
<td>12 months</td>
<td>21</td>
<td>none</td>
</tr>
<tr>
<td>AF</td>
<td>F</td>
<td>6.11</td>
<td>Connexin 26</td>
<td>birth</td>
<td>3 weeks</td>
<td>24</td>
<td>bilateral</td>
</tr>
<tr>
<td>AG</td>
<td>M</td>
<td>8.9</td>
<td>CCMV</td>
<td>birth</td>
<td>7 months</td>
<td>48</td>
<td>HA</td>
</tr>
<tr>
<td>AH</td>
<td>F</td>
<td>9.8</td>
<td>Unknown</td>
<td>birth</td>
<td>3 months</td>
<td>108</td>
<td>HA</td>
</tr>
<tr>
<td>AI</td>
<td>F</td>
<td>9.7</td>
<td>Connexin 26</td>
<td>3 months</td>
<td>3 months</td>
<td>12</td>
<td>bilateral</td>
</tr>
<tr>
<td>AJ</td>
<td>M</td>
<td>6.8</td>
<td>Unknown</td>
<td>30 months</td>
<td>32 months</td>
<td>44</td>
<td>HA</td>
</tr>
<tr>
<td>AK</td>
<td>M</td>
<td>6.1</td>
<td>Unknown</td>
<td>3 months</td>
<td>6 months</td>
<td>12</td>
<td>bilateral</td>
</tr>
<tr>
<td>AM</td>
<td>F</td>
<td>8.6</td>
<td>Connexin 26</td>
<td>birth</td>
<td>6 weeks</td>
<td>9.5</td>
<td>bilateral</td>
</tr>
</tbody>
</table>

| Mean  | 7.862 |
| SD    | 1.389 |

Stimuli

All auditory and visual recordings were made in a sound-attenuated booth, over multiple sessions. Visual objects were created and manipulated from published materials (Geisel and Geisel, 1954; 1958; Storkel, 2001). Storybook characters were created with Adobe Photoshop CC 2017 and Adobe Illustrator CC2017 (version 21).

Familiarization Stimuli

Participants completed the Familiarization task first, prior to the first Learning task. Participants did not complete Familiarization prior to the second task presentation because they were already familiarized to the task. Eight familiar objects (apple, book, box, car, flower, fork, shoe, and train) that were appropriate for children younger than the recruited age group were chosen, to try to ensure familiarity with each object. Four objects were presented on the screen, in a two-by-two grid, and participants were instructed to select one from the prompt, “Where’s the [target word]?” Once participants made a selection, feedback was provided.

Familiarization Feedback

If the selection was correct, the object appeared with a green outline to indicate a correct response. If an incorrect selection was made, the selected object was removed from the screen and participants were prompted to select another object with the prompt, “Try again!” If the participant made a correct selection, the object appeared with a green outline to indicate a correct response. If the participant made a second incorrect selection, the selected object appeared with a red outline and was removed from the screen. Participants were directed to the correct object with the prompt, “Here it is!” As the speaker stated “Here it is!” a green box appeared around the correct object. No participants made an incorrect selection during the Familiarization phase, so the experimenter asked each participant to intentionally select an incorrect object to become familiar with the Test task and feedback process. No participants reported unfamiliarity with any of the object-label pairs. Object-label pairs presented in the Familiarization phase are shown in Table 3-2.

Familiar Speech Stimuli

All speech stimuli were recorded by the same female speaker, a trained speech-language pathologist, who was native to the southeastern region of the United States. All recordings were made in a sound-treated booth. The speaker produced three tokens of each familiar object label as well as the carrier phrases, “Where’s the [target],” “Try again,” and “Here it is!” and the middle tokens were used as stimuli. The speaker sat approximately one meter (3.28 ft) from a Canon XA10 HD video camera, used for visual recordings.
### Table 3-2. Example familiar object-label pairs

<table>
<thead>
<tr>
<th>Object</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="apple" /></td>
<td>apple</td>
</tr>
<tr>
<td><img src="image" alt="book" /></td>
<td>book</td>
</tr>
<tr>
<td><img src="image" alt="box" /></td>
<td>box</td>
</tr>
<tr>
<td><img src="image" alt="car" /></td>
<td>car</td>
</tr>
<tr>
<td><img src="image" alt="flower" /></td>
<td>flower</td>
</tr>
<tr>
<td><img src="image" alt="fork" /></td>
<td>fork</td>
</tr>
<tr>
<td><img src="image" alt="shoe" /></td>
<td>shoe</td>
</tr>
<tr>
<td><img src="image" alt="train" /></td>
<td>train</td>
</tr>
</tbody>
</table>

Notes: Objects are reprinted with open access permission from [https://www.pexels.com](https://www.pexels.com) (Maeder, 2016) and [https://pixabay.com/](https://pixabay.com/) (DarkmoonArt_de, 2018a; 2018b; F_lix, 2017; Humusak, 2014. No longer available, 2014; Silberfuchs, 2014; Stux, 2014).
Acoustic speech signals were transduced with a directional condenser microphone (RÖDE NT1-A) that was placed just outside of the video frame. Microphone signals were initially amplified by a mixer/preamplifier (Mackie VLZ3) then analog band-pass filtered ($f_{bp} = 15-22000 \text{ Hz}$) and amplified using an eight-pole Butterworth filter (Krohn-Hite Model 3384). Signals were digitized by an external sound card (MOTU Microbook II) at a sampling rate of 48 kHz and saved to disk.

**Learning Phase Stimuli**

Following Familiarization, participants moved to the Learning phase where eight nonword-object pairs were presented within a narrative that was structured like a storybook. The Learning phase was presented in a split screen format. Thus, during the AV task, the speaker was presented on the top half of the screen while the storybook page, which contained one novel object and a story character, was presented on the bottom half of the screen. During the AO task, a black box was placed over the speaker’s face, so it was not visible.

**Novel objects**

Sixteen nonword-object pairs were used and eight pairs were presented in each word learning task (AV and AO). Objects were digitally-created adaptations of objects used in published work (Geisel and Geisel, 1954; 1958; Storkel, 2001) and made to resemble children’s toys. Objects were created using Adobe Photoshop CC 2017 and Adobe Illustrator CC2017 (version 21). Care was made to ensure each object was distinct in shape and color.

**Novel object labels**

Object labels were consonant-vowel-consonant (CVC) nonwords. The initial and final consonant sounds were early- or mid-acquired sounds, using criteria outlined by Shriberg, 1993. All nonwords were “lexically easy” words, meaning they had high phonotactic probability (PP) and low neighborhood density (ND) characteristics. High PP and low ND density words have been documented as advantageous for word learning across studies (Storkel, 2001; Eisenberg et al., 2002). PP and ND characteristics for each nonword were calculated using an online calculator tailored to a child corpora (Storkel & Hoover, 2010).

**Nonword selection**

Nonwords were chosen from a database of CVC nonwords (Storkel, 2013). From the database, nonwords which were classified as early- or mid-acquired (Shriberg, 1993) were extracted. From the extracted set, the mean and standard deviation (SD) of the PP and ND of the segment sum were analyzed. Nonwords with a PP of +1SD above the mean of the extracted set and a ND score -1SD from the mean were isolated. From the isolated set of high PP/lower ND words, visemes were coded for each nonword.
Visemes and visual neighborhood density

The term viseme refers to a group of sounds that, when produced, look visually identical. For instance, /b/ and /m/ are visually indistinguishable and are therefore in the same viseme group (Tye-Murray et al., 2007). Since the nonwords were CVC in structure, each nonword had a three digit viseme code. For example, the nonsense word “fob” would be coded as “2-V-1” as /f/ is in viseme group 2, all vowels are grouped into one viseme category, and /b/ is in viseme group 1. Once all nonwords were coded for visemes, they were grouped by viseme categories. From these categories, pairs of words were randomly selected, so across both stories (AV and AO), viseme category was balanced. Care was taken to represent as many viseme categories as possible. Nonwords with corresponding lexical and visual characteristics are displayed in Table 3-3.

Narrator videos

The speaker made multiple productions of each story line. Audio and video input were recorded through separate channels then temporally-aligned and combined offline with Adobe Premiere Pro.

Narratives

Each narrative was adapted from Storkel, 2001. Narratives were composed only of simple sentences as children with CIs have demonstrated difficulty processing complex sentence structure. Only one novel word was presented within a sentence and sentence structure was balanced across AV and AO narratives.

Audio recordings. Audio recordings were made in a sound-attenuated booth. The speaker read the story narrative from a custom-built digital cue display which interfaced with the free tablet application, Simple Teleprompter. As mentioned prior, audio input was transduced with a directional condenser microphone (RØDE NT1-A) that was placed just outside of the video frame. Microphone signals were initially amplified by a mixer/preamplifier (Mackie VLZ3) then analog band-pass filtered (fbp = 15-22000 Hz) and amplified using an eight-pole Butterworth filter (Krohn-Hite Model 3384). Signals were digitized by an external sound card (MOTU Microbook II) at a sampling rate of 48 kHz and saved.

Video recordings. Audio and video recordings were made simultaneously, but through different recording equipment. Audio and video recordings were aligned after recordings were made. The speaker was recorded on a black background and wore a white shirt. The speaker wore her hair in a ponytail, minimal makeup, and no visible jewelry to minimize distractions. Two iLED312-v2 flood bicolor lights were set up around the speaker, out of frame. Video recordings were made with a Canon XA10 HD video camera set up behind the digital cue display, approximately two meters (about 6.5 feet) from the speaker.
### Table 3-3. Novel object-label pairs and corresponding lexical and visual characteristics

<table>
<thead>
<tr>
<th>Nonword</th>
<th>Object</th>
<th>PP of the Segment Sum</th>
<th>ND</th>
<th>Viseme Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Auditory-Only</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yane</td>
<td><img src="yane.png" alt="Image" /></td>
<td>.16</td>
<td>8</td>
<td>7-V-7</td>
</tr>
<tr>
<td>moin</td>
<td><img src="moin.png" alt="Image" /></td>
<td>.1609</td>
<td>8</td>
<td>1-V-7</td>
</tr>
<tr>
<td>hown</td>
<td><img src="hown.png" alt="Image" /></td>
<td>.1736</td>
<td>7</td>
<td>8-V-7</td>
</tr>
<tr>
<td>dav</td>
<td><img src="dav.png" alt="Image" /></td>
<td>.1615</td>
<td>9</td>
<td>7-V-2</td>
</tr>
<tr>
<td>pome</td>
<td><img src="pome.png" alt="Image" /></td>
<td>.1695</td>
<td>9</td>
<td>1-V-1</td>
</tr>
<tr>
<td>fom</td>
<td><img src="fom.png" alt="Image" /></td>
<td>.1642</td>
<td>7</td>
<td>2-V-1</td>
</tr>
<tr>
<td>pev</td>
<td><img src="pev.png" alt="Image" /></td>
<td>.1613</td>
<td>6</td>
<td>1-V-2</td>
</tr>
<tr>
<td>kidge</td>
<td><img src="kidge.png" alt="Image" /></td>
<td>.1827</td>
<td>8</td>
<td>7-V-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Auditory-Visual</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kook</td>
<td><img src="kook.png" alt="Image" /></td>
<td>.1711</td>
<td>9</td>
<td>7-V-7</td>
</tr>
<tr>
<td>poin</td>
<td><img src="poin.png" alt="Image" /></td>
<td>.1771</td>
<td>8</td>
<td>1-V-7</td>
</tr>
<tr>
<td>hun</td>
<td><img src="hun.png" alt="Image" /></td>
<td>.1765</td>
<td>3</td>
<td>8-V-7</td>
</tr>
<tr>
<td>kov</td>
<td><img src="kov.png" alt="Image" /></td>
<td>.1664</td>
<td>6</td>
<td>7-V-2</td>
</tr>
<tr>
<td>mem</td>
<td><img src="mem.png" alt="Image" /></td>
<td>.1683</td>
<td>8</td>
<td>1-V-1</td>
</tr>
<tr>
<td>fep</td>
<td><img src="fep.png" alt="Image" /></td>
<td>.1572</td>
<td>5</td>
<td>2-V-1</td>
</tr>
<tr>
<td>bif</td>
<td><img src="bif.png" alt="Image" /></td>
<td>.1925</td>
<td>9</td>
<td>1-V-2</td>
</tr>
<tr>
<td>didge</td>
<td><img src="didge.png" alt="Image" /></td>
<td>.1603</td>
<td>9</td>
<td>7-V-4</td>
</tr>
</tbody>
</table>

CI-Simulated Speech.

Half of NH participants (N = 12) completed the learning tasks with CI-simulated speech (NH-CISim). Incorporating this group of NH participants allowed for the investigation of the influence of spectrally degraded speech input on novel word learning. Both Learning tasks (AV and AO) as well as the 4AFC task were presented with CI-simulated speech to the NH-CISim participant group. The signal was manipulated to best mimic the input of an eight-channel CI. An eight-channel simulation was selected to best represent the current constraints of CI processing. To achieve CI-simulated speech, the speech signal was divided into eight logarithmically-spaced frequency bands. Then, the amplitude envelope of each frequency band was extracted. Next, the extracted envelope information was used to modulate noise for each respective frequency band. Finally, each noise band was combined to achieve the CI-simulated speech signal. Eight-channel noise-vocoded speech was simulated with freely available online software, AngelSim (TigerCIS).

Test Phase Stimuli

The Test Phase was presented following each block of the Learning phase. The Test Phase was presented as a 4AFC task where four of the novel objects were presented in a two-by-two grid on the bottom half of the screen. For the AO task, the Test phase was presented with the black box over the speaker and for the AV task, the speaker’s face was visible. Feedback was given following each response, per the feedback procedure outlined in the Familiarization task.

Procedure

This study required one, three hour visit, or two, one and a half hour visits to the laboratory. AV and AO tasks were randomized across participants and all laboratory testing was completed in a sound-attenuated booth. Auditory stimuli were presented in the sound-field via two loudspeakers located in front of the participant, to the right and left of the presentation screen (approximately 345° and 15° azimuth). Auditory stimuli were presented at an average level of 65dBA and all visual stimuli were presented via a 24-inch screen located in front of the participant (0° azimuth). Participants were seated approximately 1m (3.28ft) away from the screen during all experimental testing, but participant-to-screen distance was adjusted on an individual basis to afford most accurate calibration and eye tracking.

Prior to administration of any tasks, all parents or guardians completed an informed consent form and participants who were seven years old and older completed an assent form. During testing, participants’ caregivers additionally completed the Learning, Executive Attention, and Functioning Scale (LEAF; Castellanos et al., 2018; Kronenberger et al., 2014) as well as a hearing history questionnaire.
Familiarization Phase

The first phase of the novel word learning task was Familiarization. Familiarization was presented exactly as the Test phase, but instead of nonword-object pairs, participants were presented with familiar object-label pairs. Familiarization format (AV or AO) and speech condition (NoSim or CI Sim) was consistent with each participant’s first assigned word learning task. Feedback was provided after each response.

Learning Phase

The Learning phase followed Familiarization. During the Learning phase, eye tracking was used to record eye gaze and visual fixations. The eye tracker was adjusted using a calibration procedure to individualize data collection to each participant. Then, eye gaze recordings were made simultaneously with the storybook presentation.

Eye Tracking

During Learning, a screen-mounted Tobii 1750 ProSpectrum eye tracker was used to record where participants were looking on the screen. The eye tracker was mounted to the bottom of the 24-inch screen, unobtrusive to the participant, and was used to quantify and record participant eye gaze. Eye tracking allowed for analysis of looking time to the narrator’s face (eyes and mouth), storybook page (object-to-be-learned, story character), and away from the screen. Eye gaze data was collected and time stamped at a rate of 150Hz.

Calibration

Calibration is the first step when collecting eye tracking data and starts with participants aligning their gaze in accordance with an outlined box on the middle of the screen. Screen, desk, and chair placements were moved slightly to best accommodate each participant. Once the participant’s eyes were aligned in the middle of the box, calibration was initiated. During calibration, eight filled circles appeared on the screen, one at a time at pre-determined x-y screen coordinates. Participants were instructed to stare at each circle until it ‘popped’ or disappeared. As the participant fixated on each circle, the eye tracker detected the infrared light reflected from the participant’s eyes respective to each x-y coordinate. See Figure 3-2, for an example calibration procedure.

Storybook Presentation

Following calibration, participants continued to the remainder of the Learning phase.
Figure 3-2. Example calibration procedure, the first step of the Learning phase

A-E represent each stage of the calibration process in the order presented from the Tobii 1750 ProSpectrum eye tracker. The clusters of red and green dots represent left and right eye fixations at each fixation point. White dots represent the cumulative fixations across eyes for each fixation point.
During Learning, two sets of eight novel object-label pairs were each presented within a story narrative that was formatted like a storybook. Two storybooks were presented (AV and AO) across one or two laboratory visits. In the case of two laboratory visits, one story was presented during each visit. The Learning phase spanned three blocks, and across each block, the amount of exposure to each object label increased.

Within each block, 10 story trials were presented. The first and last trials did not present novel objects, nor object labels, and only acted as transitions to keep the story going. The medial eight trials each presented one of the novel object-label pairs. At the start of each trial, the storybook page was presented across the entire screen for four seconds to allow the participant to scan the page, then the story page minimized to the bottom half of the screen, revealing the speaker’s face (AV) or a black box (AO). Between each trial a black screen appeared for 1.5 seconds, and functioned as an inter-stimulus-interval (ISI). During Learning, participants were instructed to remain looking and attending to the screen as best they could. The experimenter sat in the sound booth behind and away from each participant. If the participant continued to look away from the screen, the experimenter pointed to the screen to direct the participant’s attention.

Test Phase

Word learning was assessed during the Testing phase with the 4AFC referent-identification task, through a custom JAVA script. The narrator’s face was either visible on the top half of the screen, and moved dynamically with the auditory information (AV), or was covered with a black screen (AO) to provide continuity across Learning and Testing phases. Participants used a computer mouse, or pointed, to the screen to select the corresponding object referent and feedback was provided accordingly. At the start of each trial, four different objects were shuffled and presented to prevent place biases across each of the eight trials. Eight Test trials were presented after each Learning block, so participants had three opportunities to correctly identify each of the eight novel objects. Feedback was provided after each response just as it was during the Familiarization phase. An example of the Test phase feedback process, for the AO condition, is shown in Figure 3-3.

Participants completed a battery of standard assessments in addition to the experimenter-designed word learning tasks. The Peabody Picture Vocabulary Test, 4th Edition (PPVT-4) was administered to assess receptive vocabulary skills (Dunn & Dunn, 2007). The Comprehensive Test of Phonological Processing 2nd Edition (CTOPP-2), subtests of Elision, Blending, and Phoneme Isolation (Sound Matching for children who were six years old at the time of testing) was used to assess word and sound knowledge (Wagner et al., 2013). The Test of Nonverbal Intelligence 4th Edition (TONI-4; Brown et al., 2010) was used to assess nonverbal intelligence, and The Learning, Executive, and Attention Functioning Scale (LEAF; Castellanos et al., 2018) is a parental questionnaire used to assess cognitive skills like attention and working memory of the child.
1. First, during the 4AFC task, the speaker asked “Where’s the [target]?" 

2. Next, the participant was able to move the mouse over the objects. When the participant hovered over an object, the object expanded so the participant was aware where they were clicking on the screen. Here, the participant is hovering over the object in the upper right quartile.

3. Then, the participant made a selection. Here, the participant clicked on the object in the upper right quartile. This was an incorrect selection, so a red box appeared around the selected object.

4. Since the participant selected an incorrect object, the object was removed from the screen and the participant is instructed to “Try again!” This statement was spoken by the female speaker who also presented the story.

5. Again, the objects expand as the participant hovers over them with the mouse. Here, the participant is hovering over the object in the upper left quartile.

6. Lastly, the participant made a second selection of the object in the upper left quartile. This was an incorrect selection so the participant heard "Here it is!" This statement was again spoken by the same female speaker who presented the story. As the speaker stated “Here it is!” a green box appeared.

Figure 3-3. Example of the feedback process across one AO Test trial

The Multimodal Lexical Sentence Test for Children (MLST-C; Kirk et al., 2012) was used to assess speech recognition accuracy under AV and AO conditions. All standard assessments were given and scored according to their respective protocol guidelines.

**Standard Assessment Battery**

A standard assessment battery was used to measure vocabulary and phonological processing skills as well as nonverbal intelligence and executive functioning and audiovisual speech processing. Assessments were divided and administered across one or two laboratory visits. All assessments were administered by laboratory personnel with the exception of the LEAF (e.g., Castellanos et al., 2018) which is a parental report measure.

**The Peabody Picture Vocabulary Test, 4th Edition**

The PPVT-4 was administered to assess receptive single word vocabulary skills. Each child started the assessment at their calculated age. The child was shown a page with four pictures in a two-by-two format and presented a word. Participants were instructed to point to the picture or indicate the number for the picture that corresponded to the target word and guessing was encouraged. Participants continued identifying pictures until they hit a ceiling of eight incorrect responses in one set.

**Comprehensive Test of Phonological Processing, 2nd Edition**

The CTOPP-2 subtests of Elision, Blending, and Phoneme Isolation (or Sound Matching) were used to assess word and sound knowledge. Elision and Phoneme Isolation subtests were administered by the experimenter. The Blending subtest was presented via CD through two loudspeakers located in front of the participant (approximately 345° and 15° azimuth). For all subtests of the CTOPP-2, ceiling performance was reached and administration ceased when the participant missed three consecutive items.

**Elision Subtest**

The Elision subtest is used to assess the child’s knowledge of phonemes. The experimenter asked the child to repeat a word. Then, the child was asked to repeat the word without one or multiple sounds. For example, “Say the word planes without saying /n/.”
Blending Words Subtest

The Blending subtest was used to assess the child’s ability to combine sounds to form words. A female recorded speaker asked, “What word do these sounds make: cow-boy?” Then the child is asked to combine the two to form one word, cowboy. The subtest progresses from syllables to single phonemes (e.g., “What word do these sounds make: g-r-a-s-h-o-p-ə?”).

Phoneme Isolation Subtest

Participants, seven years old and older, completed the Phoneme Isolation subtest, which aimed to understand the child’s ability to identify specific phonemes within a provided word. The experimenter presented a word then asked the child to indicate a specific sound within that word. For example, “What is the second sound in the word island?”

Sound Matching Subtest

Six-year-old participants completed the Sound Matching subtest instead of the Phoneme Isolation subtest. Sound matching incorporates a booklet with cartoon pictures that represented test words. Participants were given the labels for different pictures then given a word and asked to indicate the picture whose label either started or ended with the same sound as the word provided.

The Test of Nonverbal Intelligence, 4th Edition

Each page of the TONI-4 contains several abstract shapes that correspond to each other on one or multiple dimensions (e.g., shape, size, etc.) and one of the shapes is missing. The child was asked to choose from a selection of four or six items which completes the abstract set of shapes. Difficulty increased as the child progressed and guessing was encouraged. Ceiling was reached when participants gave three incorrect responses within five consecutive items.

Learning Executive and Attention Functioning (LEAF) Scale

The LEAF is a parental questionnaire designed to provide information about executive function ability for each child across sub-areas such as attention, learning, and working memory. The LEAF is a survey of 55 statements, grouped into 11 subscales. Parents were asked to read each statement and circle one of four options to indicate how much they are in agreement that the statement pertains to their child. Example statements include “Can’t do more than one thing at a time” and “Has trouble sounding out new words when reading.” LEAF subscales include:
• Comprehension and conceptual learning
• Factual memory
• Attention
• Processing speed
• Visual-spatial organization
• Sustained sequential processing
• Working memory
• Novel problem solving
• Mathematics skills
• Basic reading skills
• Written expression skills

The Multimodal Lexical Sentence Test for Children (MLST-C)

In addition to the aforementioned standard assessments, the Multimodal Lexical Sentence Test for Children (MLST-C; Kirk et al., 2012) was administered in AO and AV conditions. The MLST-C is an assessment of sentence recognition abilities when sentences are presented in AO and AV formats. The MLST-C is scored online as a percent correct score. Participants were administered one list of sentences across each condition (AO and AV).

Data Recording

Learning phase stimuli were presented using a timeline, which is a stimulus presentation feature of the Tobii ProLab software. ISIs were used between story trials to provide a break between presentations. Eye tracking recordings were made across the entire timeline, but only eye gaze data from the story trials, not ISIs, were analyzed. At the start of each Learning trial, the object was displayed on the screen for four seconds to allow the participant to scan the object, then it minimized to the bottom half of the screen to display either the speaker (AV) or a black box (AO) on the top half of the screen. To afford a fair analysis of time spent looking to each area of interest (AOI) on the screen, eye gaze analyses did not start until the speaker (AV) or black box (AO) was displayed. See Figure 3-4 for timeline.

Calibration

Eight-point calibration was attempted for all participants. Calibration data were recorded for all but two participants. These two participants were unable to gaze at each calibration point long enough to collect the calibration data. After multiple attempts to calibrate, both participants became agitated, so they each completed the tasks without individualized calibration and eye tracking recordings. The two participants who were unable to complete calibration were not included in eye gaze analyses.
Figure 3-4.  Example timeline of the first part of a Learning trial, prior to the onset of auditory cues

The character and object were first presented across the entire screen, then over time, moved to the bottom 50% of the screen over four seconds, then the black box or speaker’s face appeared on the top 50% of the screen. Eye tracking started at the onset of the speaker’s face (AV) or when the black box appears (AO).
Areas of Interest

AOIs were regions of space on the screen that corresponded with a visual area of interest to this study, such as the speaker’s mouth, eyes, and the object-to-be-learned. Videos were coded frame-by-frame and each visualization shift was coded and sorted into each AOI category.

Coding Eye Gaze

All eye gaze recordings were made and stored with the Tobii ProLab software. Visualizations were coded and recorded with Microsoft Excel while muted so the experimenter was unable to hear the auditory stimuli to prevent biases. Eye visualizations were coded into the following AOIs: (1) speaker’s mouth (2) speaker’s eyes (3) another area on the speaker’s face, labeled “face other” (4) the story character (5) the object-to-be-learned. Most eye visualizations were clearly in one of the AOIs, but best guesses were made for visualizations that were not clearly within the boundaries of an AOI. Eye gaze visualizations to each AOI were summed for each trial across Learning blocks. Once the amount of time to each AOI was summed for each trial, the AOI fixation time was calculated as a proportion of time spent looking based on total trial length. For analyses of block, the proportion of time spent looking to each AOI across trials was averaged across Learning blocks.

Test Phase

All participant responses were automatically recorded and saved through the custom JAVA script then written into Microsoft Excel files. Responses were analyzed as a number or proportion correct across each Test block. First and second responses, following feedback, were analyzed individually.
CHAPTER 4. RESULTS

This study had two primary aims (1) to understand how access to AV speech information impacts novel word learning success in children with CIs and their NH peers, and (2) to characterize individual looking patterns as they relate to individual word learning outcomes for children with CIs. To address these aims, children with CIs and children with NH completed two novel word learning tasks (AO and AV). While participants completed these tasks, eye tracking was implemented to record visual fixations to different AOIs on the presentation screen. In addition to the experimenter-designed learning tasks, participants also completed a battery of standard assessments. These tasks and assessments were used to address three research questions (1) How does access to AV information impact novel word learning success for children with CIs and children with NH listening to non-simulated and CI-simulated speech? (2) How do individual patterns of visual attention during learning relate to individual word learning outcomes for children with CIs and children with NH? (3) What measured factors (e.g., hearing history and demographic information) contribute to novel word learning across AV and AO tasks?

How Does Access to AV Information Impact Novel Word Learning Success for Children with CIs and Children with NH Listening to Non-Simulated and CI-Simulated Speech?

Each participant group (CI, NH-NoSim, and NH-CISim) completed two novel word learning tasks (AV and AO). Word learning was assessed during the Test phase, via a 4AFC referent identification task, presented after each of three blocks in the Learning phase (refer to Figure 3-1 for task outline). Number of words correct across each block (1, 2, 3) for each word learning task (AV, AO) was scored for each participant group (CI, NH-NoSim, NH-CISim) and analyzed with a repeated measures analysis of variance (ANOVA), with a Bonferroni correction applied. In the ANOVA, Block (1, 2, 3) and task type (AV, AO) were included as within-subjects variables and group (CI, NH-NoSim, NH-CISim) was included as a between-subjects variable. Correct responses were scored as number correct, out of a possible eight, across each of the three blocks. Thus, by the end of Block three, participants could have achieved a maximum of 24 correct selections within each task (AV and AO). Results from the first object selection were included in the analyses.

Participants demonstrated similar word learning outcomes across AV and AO tasks, with no significant difference across group, $F(1, 33) = 2.358, p = .110, \eta^2_p = .125$. There was also no main effect of Task, $F(1, 33) = 0, p = 1, \eta^2_p = .00_0$, but the ANOVA did reveal a task by group interaction, $F(2, 66) = 3.608, p = .038, \eta^2_p = .179$. Although all groups performed similarly across the AV and AO tasks, not all groups demonstrated higher performance in the AV task, as predicted. Children with CIs and children in the NH-NoSim group demonstrated higher performance in the AV task as compared to the AO task, but children in the NH-CISim group demonstrated higher performance in the
AO task as compared to the AV task, shown in Figure 4-1. To better understand the task by group interaction, a one-way ANOVA was conducted for task type. There was not a significant different between AV task performance, $F(2,35) = 3.086, p = .059$ or AO task performance, $F(2,35) = 1.812, p = .179$, across groups. Paired t-tests were conducted to compare task type for each group. A significant difference between task types was found for participants in the NH-CISim group ($t(11) = -2.424, p = .033$) indicating significantly more words were learned in the AO task ($M = 12.50, SD = 3.09$) as compared to the AV task ($M = 10.75, SD = 3.911$).

Word Learning Across Blocks

Learning was assessed after each block of the Learning phase with a 4AFC referent identification task. As participants completed each Learning block, it was anticipated that performance would improve. Across each Learning block, the amount of exposure to each object label increased accordingly, so additional experience with each nonword-object label, compounded with feedback given after each block during the 4AFC task (Test phase), was expected to support a greater number of words learned over time. The results of the ANOVA showed a significant main effect of Block, $F(1.823, 60.175) = 11.537, p < .01$, $\eta_p^2 = .259$, thus, participants demonstrated improved word learning across each task in accordance with greater exposure and feedback. Paired samples t-tests were used to investigate differences within a task, across blocks, as well as, within a block, across tasks. All groups displayed a significant increase in number of words learned by Block three in the AV task. Only children in the NH-CISim group displayed significant learning over time in the AO task, $t(11) = -3.96, p = .002$. Word learning outcomes across blocks is shown in Figure 4-2.

Item Analyses

Item analyses were conducted to ensure that some words were not significantly easier or harder to learn over time within each task. A linear mixed-model with a UN_CS covariance structure was used to account for the repeated measures of block and word and the Kronecker product was applied. For the AV task, a significant main effect of group, $F (2, 83.224) = 3.481, p = .037$ and block, $F(2, 76.085) = 8.111, p = .001$ were found with no significant interactions and no main effect of word, $F (7, 244.582) = 1.893, p = .071$. For the AO task, a significant main effect of block, $F (2, 77.983) = 3.398, p = .038$ was found as well as a block by word interaction, $F (14, 384.051) = 2.145, p = .009$. Pairwise comparisons were used to investigate the block by word interaction found in the AO task. Performance across five of the nonwords in the AO task did not significantly improve across the AO task, but performance across three of the nonwords (‘fom,’ ‘hown,’ and ‘yane’) did significantly change over time, where performance for each nonword significantly improved by Block three [‘fom’ Block one to Block three, $t(2) = 3.39, p = .001$] (‘hown’ Block one to Block three, $t(2) = 3.088, p = .002$) (‘yane’ Block two to Block three, $t(2) = 2.467, p = .015$).
Figure 4-1. Mean number of words learned across AV and AO conditions for each participant group

Mean number correct was averaged across all blocks. The mean and standard error are indicated with black filled circles and error bars for each group. Participants in the CI group are indicated with a green dashed line, participants in the NH-NoSim group are indicated with a blue dotted line, and participants in the NH-CISim group are indicated with a solid burgundy line.
Figure 4-2. Mean number of words learned across AV and AO conditions for each participant, by block

AV outcomes are shown with blue bars, AO outcomes are shown with green bars. Significant differences are indicated with brackets. Bracket color corresponds to the task in which the difference was found. The burgundy horizontal bracket indicates a significant difference between tasks in Block three for the NH-NoSim group.
How Do Patterns of Visual Attention During Learning Relate to Individual Word Learning Outcomes?

The second aim of this study was to characterize individual looking patterns as they relate to individual word learning outcomes for children with CIs. Visual fixations were analyzed as a proportion of time spent looking to multiple AOIs across blocks within the AV word learning task. Visual fixations were not analyzed for the AO word learning task because the speaker’s face was not present, so comparisons of fixations across AV and AO tasks was not balanced. Two children in the CI group were unable to complete calibration, and eye gaze data collected across the task was less than 20% of the total Learning phase, so those two participants (AD and AK) were not included in the analyses. A repeated-measures ANOVA was conducted with Block (1,2,3) and AOI (speaker’s eyes, speaker’s mouth, face other, object-to-be-learned, and character) as within-subjects variables and group (CI, NH-NoSim, NH-CISim) as the between-subjects factor, and a Bonferroni correction was applied. Group was included in the first analysis to compare visual attention patterns of the CI group to the NH participant groups. There was a main effect of AOI, \( F(1.864, 57.239) = 61.688, p < .01, \eta^2_p = 1.00 \), and a Block by AOI interaction, \( F(3.5847, 111.114) = 6.545, p < .01, \eta^2_p = .928 \), but the ANOVA revealed no main effect of group, \( F(2, 31) = .589, p = .561, \eta^2_p = .139 \).

Areas of Interest

AOIs included the speaker’s face, the speaker’s eyes, “face other” which represented visualizations to the speaker’s face other than the eyes and mouth, the story character, and the novel object. Participants spent a longer proportion of the trial looking to the speaker’s mouth (\( M = .523 \)) than any other AOI [(character, \( M = .119 \)), (object, \( M = .151 \)), (face other, \( M = .022 \)), (eyes, \( M = .149 \)]. Paired-samples t-tests were used to investigate the Block by AOI interaction, first with participant groups collapsed as there was no main effect of group. The t-tests revealed multiple differences in looking across blocks. The amount of time spent looking to the character decreased from Block one to Block two, \( t(33) = 2.262, p = .03 \), and from Block two to Block three, \( t(33) = 2.53, p = .005 \). The amount of time spent looking at the object also decreased from Block one to Block two, \( t(33) = 2.076, p = .046 \), and from Block two to Block three, \( t(33) = 4.761, p < .01 \). The amount of time spent looking at the speaker’s face to somewhere apart from the eyes and mouth (“face other”) did not change across block. The amount of time spent looking to the speaker’s mouth increased from Block one to Block two, \( t(33) = -2.04, p = .049 \), and from Block two to Block three, \( t(33) = -3.037, p = .005 \). Looking time to the speaker’s mouth did not significantly change across time, but participants spent a significantly longer proportion of the trial attending to the speaker’s mouth within the task. Across the first block, the proportion of time spent looking to the speaker’s mouth was significantly greater than all other AOIs [(character, \( t(33) = 6.778, p < .01 \)), (object, \( t(33) = 6.220, p < .01 \)), (face other, \( t(33) = 12.692, p < .01 \)), (eyes, \( t(33) = 7.472, p < .01 \))]. Across the second block, the proportion of time spent looking to the speaker’s mouth was significantly greater than all other AOIs [(character, \( t(33) = 8.616, p < .01 \)), (object, \( t(33) = 7.156, p < .01 \)), (face other, \( t(33) = 12.784, p < .01 \)), (eyes, \( t(33) = 6.229 \)), (character, \( t(33) = 8.616, p < .01 \)), (object, \( t(33) = 7.156, p < .01 \)), (face other, \( t(33) = 12.784, p < .01 \)), (eyes, \( t(33) = 6.229 \))].
Across the third block, the proportion of time spent looking to the speaker’s mouth was significantly greater than all other AOIs [(character, \(t(33) = 8.481, p < .01\)), (object, \(t(33) = 7.952, p < .01\)), (face other, \(t(33) = 11.639, p < .01\)), (eyes, \(t(33) = 4.925, p < .01\))]. The proportion of time spent looking to each AOI across participant groups is shown in Figure 4-3.

The CI group is the only group that significantly altered their looking behavior to the speaker’s mouth over time. Neither NH group changed their looking time to the speaker’s mouth from Block one to Block three, \(t(9) = -2.706, p = .024\). Not only did the looking time to the speaker’s mouth change across blocks, but it was highly correlated with word learning outcomes by the end of the task, \(r(10) = .699, p = .024\). Mean proportion of time spent looking to each AOI across time, as a function of word learning, is plotted for each CI participant in Figures 4-4 through 4-6. The NH-CISim group is the only group that significantly changed their looking behavior to the speaker’s eyes over time. Looking time to the speaker’s eyes significantly increased from Block one to Block three, \(t(11) = -2.634, p = .024\).

**Individual Differences in the CI Group**

The aim of the second research question was not only to quantify looking to each AOI, but to investigate the relationship of different visualizations to individual word learning outcomes, specifically for the CI group. Large individual differences are common among children with CIs on measures of speech perception and learning, and this trend of large variability was apparent on learning measures in this study as well. As shown in Figure 4-7 many children with CIs displayed learning across AV and AO tasks below 50% (12 words) correct, but many children with CIs displayed learning above 50% (12 words) correct. As has been documented in prior work (e.g., Eisenberg et al., 2002; Kirk et al., 2007), children in the CI group were divided into “poorer” and “higher” performers, based on their AV learning outcomes to better characterize individual performance.

K-means clustering was used to determine the ‘higher’ and ‘poorer’ participant clusters. The two clusters performed significantly different on each word learning task [AV, \(F(1,11) = 13.846, p = .004\); AO, \(F(1,11) = 40, p < .01\)]. To investigate if the differences across these performance clusters were correlated with looking to AOIs, correlations were conducted for looking to the speaker’s mouth and word learning performance, for each performance cluster. Outcomes are displayed in Figure 4-8.
Figure 4-3. Proportion of time spent looking to each AOI across blocks for each participant group

Looking for Block 1 is shown with a green line, Block 2 is shown with a blue line, and Block 3 is shown with a pink line. Mean time and standard error are shown with a circle or triangle and error bars at each AOI.
Figure 4-4. Correlation between proportion of time spent looking to each AOI and proportion of words correct across Block 1 for CI participants

Each CI participant is indicated with their subject code
Figure 4-5. Correlation between the proportion of time spent looking to each AOI and proportion of words correct across Block 2 for CI participants.

Each CI participant is indicated with their subject code.
Figure 4-6. Correlation between the proportion of time spent looking to each AOI and proportion of words correct across Block 3 for CI participants

Each CI participant is indicated with their subject code. The burgundy box represents a significant correlation.
Figure 4-7. Correlation among number of words learned on the AV and AO tasks for children in the CI group

Participants are indicated with their participant identification code
Figure 4-8. Correlations between the proportion of time spent looking to the speaker’s mouth and number of words learned in the AV task

Correlations are shown for each block. Within each block, separate correlations are plotted for higher and poorer performers. Higher performers are indicated with green text and a green regression line. Poorer performers are indicated with blue text and a blue regression line.
Participant Feedback Rate

Participants who made an incorrect selection following the first presentation within a Test trial were given the opportunity to select a second object during the feedback process. Second object selections among the CI group were analyzed to assess individual performance by second selections needed across participants. It was suspected that participants who were poorer performers needed a greater rate of feedback than higher performers. Participant feedback ‘rate’ refers to the necessity of providing feedback as well as the success of providing feedback after each Test trial. Feedback rate was coded where ‘3’ indicated that within that trial, the participant made a correct selection during their first attempt and no feedback was required, ‘2’ indicated that within that trial, the participant made an incorrect selection the first time, then once that object was removed, made a correct second selection, and ‘1’ indicated that the participant failed to make a correct selection after two attempts within that trial. Feedback rate was averaged across trials for each block. A linear mixed model with a Diagonal covariance type was used to investigate the feedback rate for each participant in the CI group. Feedback rate for CI participants, across AV and AO tasks, is displayed in Figure 4-9 and Figure 4-10, respectively.

A one-way ANOVA was used to investigate the relationship between CI performance group and feedback rate, across block within the AV task. Performance groups significantly differ in Block three where higher performers selected the correct object without feedback more often than children in the poorer performers group, $F(1,11) = 6.241, p = .032$. A second one-way ANOVA was conducted to investigate the same relationship in the AO task. Performance groups significantly differed in Block two of the task, where higher performers needed less feedback than poorer performers, across the second block of trials $F(1,11) = 17.286, p = .002$.

What Factors Contribute to Novel Word Learning Success Across AV and AO Tasks?

The third aim of this study was to understand what measured factors were correlated with novel word learning success on the given task. Correlations were conducted for hearing history and demographic variables, word learning outcomes, and standardized scores from the assessment battery.

Variables

Hearing history variables were only included for children with CIs. Demographic variables were included for all groups. Correlations were conducted for each group as they pertain to outcomes. Correlation variables are listed below:
Figure 4-9. Feedback rate across CI participants for each Block in the AV task

Bars with diagonal stripes represent outcomes across poorer performers and solid bars represent outcomes across higher performers. Error bars represent standard error.
Figure 4-10. Feedback rate across CI participants for each Block in the AO task

Bars with diagonal stripes represent outcomes across poorer performers and solid bars represent outcomes across higher performers. Error bars represent standard error.
**Hearing History, Device, Aural (Re)Habilitation Variables – CI Only**

- Age of amplification and implantation
- Unilateral/bilateral implantation
- Amount of time spent in speech therapy per week

**Demographic and Test Variables – All Participants**

- Maternal education level
- PPVT-4 standard score
- CTOPP-2 scaled scores across subtests (Elision, Blending Words, Phoneme Isolation/Sound Matching) TONI-4 index score
- LEAF summed scores across each of the 11 subscales
- MLST-C scores across AV and AO conditions

**CI Group Correlations**

The CI group correlations were used to investigate any factors related to hearing loss and amplification that may be contributing to the observed word learning outcomes. All CI participants were grouped for the first analyses. Better learning outcomes in the AO task were significantly associated with better outcomes in the AV task, \( r(12) = .71, p = .009 \), bilateral implantation, \( r(12) = .894, p < .01 \), and hearing aid use, \( r(12) = -.701, p = .011 \). Performance on the AO task was also significantly correlated with age of amplification, \( r(12) = -.739, p = .006 \), as shown in Figure 4-11, performance on the CTOPP-2 Blending subtest, \( r(12) = .684, p = .014 \) (Figure 4-12), the LEAF Attention Subscale score, \( r(12) = -.714, p = .009 \) (Figure 4-13), the LEAF Sustained Sequential Processing Subscale score, \( r(12) = -.626, p = .029 \) (Figure 4-14), the LEAF Working Memory Subscale score, \( r(12) = -.702, p = .011 \) (Figure 4-15), the LEAF Novel Problem Solving Subtest score, \( r(12) = -.579, p = .049 \) (Figure 4-16), and the LEAF Mathematics Skills Subtest score, \( r(12) = -.608, p = .036 \) (Figure 4-17). No significant correlations were found between word learning task and age of implantation or amount of time spent participating in speech therapy when all CI participants are grouped. Performance on the AV task was significantly correlated with outcome on the TONI, \( r(12) = .621, p = .031 \) (Figure 4-18). Performance on both AV and AO tasks was significantly correlated with the LEAF Factual Memory Subscale score [AV; \( r(12) = -.653, p = .021 \); AO; \( r(12) = -.651, p = .022 \) (Figure 4-19)], the LEAF Processing Speed Subscale score [AV; \( r(12) = -.711, p = .009 \); AO; \( r(12) = -.607, p = .036 \) (Figure 4-20)], and the LEAF Visual-Spatial Organization Subscale score [AV; \( r(12) = -.604, p = .038 \); AO; \( r(12) = -.81, p = .001 \) (Figure 4-21)]. It is important to note that the LEAF is arranged so that a lower score is representative of better executive functioning, attention, and working memory skills. Learning outcomes were not correlated with scores from the MLST-C, which was a task of sentence recognition, presented in AV and AO formats.
Figure 4-11. Correlation between word learning outcomes on the AO task and age of amplification

Participants are indicated with their participant identification code.

Figure 4-12. Correlation between word learning outcome on the AO task and scaled score on the CTOPP-2, Blending Subtest

Participants are indicated with their participant identification code.
Figure 4-13. Correlation between word learning outcome on the AO task and summed score on the LEAF Attention subscale

Participants are indicated with their participant identification code.

Figure 4-14. Correlation between word learning outcome on the AO task and summed score on the LEAF Sustained Sequential Processing subscale

Participants are indicated with their participant identification code.
Figure 4-15. Correlation between word learning outcome on the AO task and summed score on the LEAF Working Memory subscale

Participants are indicated with their participant identification code.

Figure 4-16. Correlation between word learning outcome on the AO task and summed score on the LEAF Novel Problem Solving subscale

Participants are indicated with their participant identification code.
Figure 4-17. Correlation between word learning outcome on the AO task and summed score on the LEAF Mathematics Skills subscale

Participants are indicated with their participant identification code.

Figure 4-18. Correlation between word learning outcome on the AV task and TONI index score

Participants are indicated with their participant identification code.
Figure 4-19. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on LEAF Factual Memory subscale

Participants and outcomes for the AV task are indicated with blue text and participants with outcomes for the AO task are indicated with green text. Corresponding regression lines are also displayed in blue or green.
Figure 4-20. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on LEAF Processing Speed subscale

Participants and outcomes for the AV task are indicated with blue text and participants with outcomes for the AO task are indicated with green text. Corresponding regression lines are also displayed in blue or green.
Figure 4-21. Correlation between word learning outcomes on the AV and AO tasks and parental report on summed score on the LEAF Visual-Spatial Organization subscale

Participants and outcomes for the AV task are indicated with blue text and participants with outcomes for the AO task are indicated with green text. Corresponding regression lines are also displayed in blue or green.
Characterizing ‘High’ and ‘Poor’ Performance

Age of amplification, age of implantation, and phonological processing skills, as measured via the CTOPP-2 Blending subtest, differentiated the ‘higher’ and ‘poorer’ performers within the CI group using an independent samples t-test with equal variances assumed. The higher performers were implanted earlier in life, $t(10) = -2.899, p = .016$, received amplification earlier in life, $t(10) = -2.293, p = .045$, and demonstrated better phonological processing on the Blending subtest of the CTOPP-2, $t(10) = 3.588, p = .005$.

Correlations with NH Groups

Demographic and test variables were included in correlations with AV and AO learning outcomes. Correlations were separated across NH-CISim and NH-NoSim groups. Correlations are shown for the AV task in Table 4-1 and for the AO task in Table 4-2, across each participant group.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Participant Group (AV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CI</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>$r(12) = .233, p = .465$</td>
</tr>
<tr>
<td>CTOPP-2: Elision</td>
<td>$r(12) = .296, p = .351$</td>
</tr>
<tr>
<td>CTOPP-2: Blending</td>
<td>$r(12) = .563, p = .351$</td>
</tr>
<tr>
<td>CTOPP-2: Phoneme Isolation</td>
<td>$r(12) = .392, p = .208$</td>
</tr>
<tr>
<td>TONI-4</td>
<td>$r(12) = .621, p = .031$</td>
</tr>
<tr>
<td>LEAF: Conceptual Learning</td>
<td>$r(12) = .353, p = .261$</td>
</tr>
<tr>
<td>LEAF: Factual Memory</td>
<td>$r(12) = -.653, p = .021$</td>
</tr>
<tr>
<td>LEAF: Attention</td>
<td>$r(12) = -.504, p = .095$</td>
</tr>
<tr>
<td>LEAF: Processing Speed</td>
<td>$r(12) = -.711, p = .009$</td>
</tr>
<tr>
<td>LEAF: VS Organization</td>
<td>$r(12) = .604, p = .038$</td>
</tr>
<tr>
<td>LEAF: SS Processing</td>
<td>$r(12) = -.485, p = .110$</td>
</tr>
<tr>
<td>LEAF: Working Memory</td>
<td>$r(12) = -.483, p = .111$</td>
</tr>
<tr>
<td>LEAF: Novel Problem-Solving</td>
<td>$r(12) = .256, p = .422$</td>
</tr>
<tr>
<td>LEAF: Mathematics Skills</td>
<td>$r(12) = -.254, p = .426$</td>
</tr>
<tr>
<td>LEAF: Basic Reading Skills</td>
<td>$r(12) = -.407, p = .189$</td>
</tr>
<tr>
<td>LEAF: Written Expression Skills</td>
<td>$r(12) = -.533, p = .074$</td>
</tr>
<tr>
<td>MLSTC-AO</td>
<td>$r(12) = .342, p = .277$</td>
</tr>
<tr>
<td>MLSTC-AV</td>
<td>$r(12) = .192, p = .551$</td>
</tr>
</tbody>
</table>

Note: Significant correlations are indicated with bold text. VS, Visual-Spatial. SS, Sustained Sequential.
Table 4-2. Correlation between demographic variables and AO learning outcomes across participant groups

<table>
<thead>
<tr>
<th>Measure</th>
<th>Participant Group (AO)</th>
<th>CI</th>
<th>NH-CISim</th>
<th>NH-NoSim</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-4</td>
<td></td>
<td>$r(12) = .089, p = .783$</td>
<td>$r(12) = -.058, p = .857$</td>
<td>$r(12) = .324, p = .305$</td>
</tr>
<tr>
<td>CTOPP-2: Elision</td>
<td>$r(12) = .368, p = .240$</td>
<td>$r(12) = -.590, p = .043$</td>
<td>$r(12) = .367, p = .240$</td>
<td></td>
</tr>
<tr>
<td>CTOPP-2: Blending</td>
<td>$r(12) = .684, p = .014$</td>
<td>$r(12) = 110, p = .735$</td>
<td>$r(12) = -.202, p = .529$</td>
<td></td>
</tr>
<tr>
<td>CTOPP-2: Phoneme Isolation</td>
<td>$r(12) = .263, p = .409$</td>
<td>$r(12) = -.312, p = .323$</td>
<td>$r(12) = .233, p = .465$</td>
<td></td>
</tr>
<tr>
<td>TONI-4</td>
<td>$r(12) = .342, p = .276$</td>
<td>$r(12) = -.586, p = .045$</td>
<td>$r(12) = .455, p = .137$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Conceptual Learning</td>
<td>$r(12) = -.355, p = .258$</td>
<td>$r(12) = -.208, p = .516$</td>
<td>$r(12) = .010, p = .976$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Factual Memory</td>
<td>$r(12) = -.651, p = .022$</td>
<td>$r(12) = .264, p = .407$</td>
<td>$r(12) = -.028, p = .930$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Attention</td>
<td>$r(12) = -.714, p = .009$</td>
<td>$r(12) = -.104, p = .748$</td>
<td>$r(12) = -.068, p = .835$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Processing Speed</td>
<td>$r(12) = -.607, p = .036$</td>
<td>$r(12) = -.196, p = .542$</td>
<td>$r(12) = -.084, p = .795$</td>
<td></td>
</tr>
<tr>
<td>LEAF: VS Organization</td>
<td>$r(12) = -.81, p = .001$</td>
<td>$r(12) = .720, p = .008$</td>
<td>$r(12) = -.077, p = .811$</td>
<td></td>
</tr>
<tr>
<td>LEAF: SS Processing</td>
<td>$r(12) = -.626, p = .029$</td>
<td>$r(12) = -.060, p = .852$</td>
<td>$r(12) = -.292, p = .357$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Working Memory</td>
<td>$r(12) = -.702, p = .011$</td>
<td>$r(12) = .181, p = .573$</td>
<td>$r(12) = -.064, p = .843$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Novel Problem-Solving</td>
<td>$r(12) = -.579, p = .049$</td>
<td>$r(12) = .584, p = .046$</td>
<td>$r(12) = -.003, p = .993$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Mathematics Skills</td>
<td>$r(12) = -.608, p = .036$</td>
<td>$r(12) = -.230, p = .473$</td>
<td>$r(12) = -.343, p = .274$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Basic Reading Skills</td>
<td>$r(12) = -.261, p = .413$</td>
<td>$r(12) = -.046, p = .887$</td>
<td>$r(12) = -.229, p = .474$</td>
<td></td>
</tr>
<tr>
<td>LEAF: Written Expression Skills</td>
<td>$r(12) = -.344, p = .274$</td>
<td>$r(12) = -.049, p = .880$</td>
<td>$r(12) = -.155, p = .631$</td>
<td></td>
</tr>
<tr>
<td>MLSTC-AO</td>
<td>$r(12) = .599, p = .059$</td>
<td>$r(12) = .316, p = .318$</td>
<td>$r(12) = .307, p = .332$</td>
<td></td>
</tr>
<tr>
<td>MLSTC-AV</td>
<td>$r(12) = .498, p = .099$</td>
<td>$r(12) = .437, p = .156$</td>
<td>$r(12) = .329, p = .296$</td>
<td></td>
</tr>
</tbody>
</table>

Note: Significant correlations are indicated with bold text. VS, Visual-Spatial. SS, Sustained Sequential.
Multiple studies have reported a significant speech perception advantage when the listener has access to visual and auditory speech cues (AV), as opposed to only auditory (AO), in quiet and noisy environments (e.g., Helfer & Freyman, 2005; Schwartz et al., 2004; Sumby & Pollack, 1954). Although AV speech information has proven beneficial under the demands of a word or sentence recognition task, the benefit of AV speech information during a novel word learning task is not well understood. A learning task requires engagement of multiple peripheral and central processes. Specifically, novel word learning requires: (a) recognition of a word as novel, often embedded within a fluent speech stream; (b) association of that novel word with its referent, and (c) retention of this novel object-referent pair in short- then long-term memory where it can be recalled the next time the object or referent is encountered. This arguably requires greater cognitive resources than a speech recognition task which requires perception and recall of known information. Although AV speech information boosts perceptual outcomes for word and sentence recognition, the benefit of AV speech for novel word learning may be altered in accordance with the different cognitive demands of the task.

Many children with CIs fall behind their NH peers on vocabulary measures and also learn new words at a slower rate, but for children with CIs, vocabulary outcomes are widely variable (Lund, 2015). Often, children with CIs are included in a mainstreamed classroom and expected to learn and function like their NH peers, but the influence of visual speech cues on novel word learning has received little attention despite the fact children with CIs have access to these cues in their everyday environments. Understanding the contribution of AV speech information to word learning success will improve our understanding of novel word learning, and the influence of visual speech on word learning in this group, and could help explain some of the large individual differences reported (e.g., Geers et al., 2016; Niparko et al., 2010).

**Novel Word Learning Outcomes**

This study was guided by three primary research questions, designed to address the utility of visual speech information for novel word learning and to characterize individual patterns of visual attention as they relate to novel word learning outcomes. The first research question was, “How does access to AV speech information impact novel word learning success for children with CIs, children with NH listening to non-simulated speech, and children with NH listening to CI-simulated speech, when compared to novel word learning in an AO condition?” It was hypothesized that all participants would demonstrate an increase in novel word learning when presented with the speaker’s face (AV) as compared to only listening to the presented information (AO). Per the IRH (Bahrick & Lickliter, 2000), a signal which stimulates more than one sensory modality (i.e., AV) should recruit more attention than a signal which only stimulates one sensory modality (i.e., AO). This recruitment of attention is beneficial for listener engagement and should best support learning.
All participants demonstrated similar word learning performance across the AV task where word learning outcomes significantly improved by the third block of trials. Across each block of trials, participants received an increasing number of exposures to each object label, and after each test trial, participants received feedback after their response(s). Thus, it was anticipated that participants would achieve a greater number of words learned by the end of Block three, in accordance with the additional exposures and feedback provided. Although all participant groups demonstrated significant gains in learning across the AV task, there was no significant difference in learning outcomes by the third block of trials across the AV and AO tasks. Thus, the addition of visual speech cues did not make a poor word learner, a significantly better word learner. This also suggests that children who demonstrate difficulty learning new words may not be able to reliably capitalize on visual speech information. This outcome is consistent with a recently published study which reported no significant differences in novel word learning outcomes among CI participants who were taught novel object-referent pairs in either an AO or AV format (McDaniel et al., 2018).

Although this finding is consistent with prior work (McDaniel et al., 2018), it should be interpreted with caution. It is possible that the lack of main effect of task could be attributed to the underpowered sample of participants with CIs in this study. An a priori power analysis using G*Power 3.1 (Faul et al., 2007; Faul et al., 2009), with an alpha of .05, indicated that a sample size of at least 22 CI participants is necessary to support the current study; however, only 13 children with CIs participated, and 12 were included in the analyses. Currently, the difference in word learning outcomes across tasks, for children with CIs, is marginally significant, so additional data may help demonstrate a significant difference in learning across the tasks by the third block of trials. If a significant difference emerges for the CI group, by the end of the third block of trials, following collection of additional data, it would suggest that children with CIs are able to capitalize on visual speech information during a novel word learning task.

The lack of difference between AV and AO outcomes among the NH participants in this task may be attributed to the participant age range in the current study. Jerger and colleagues (2009) have proposed that AV integration skills regress in the early elementary school years, around approximately five-to-nine years of age. Jerger and colleagues have taken a Dynamic Systems theory (Smith & Thelen, 2003) approach when describing this trajectory and suggest that AV integration skills are not lost during this time period, but instead regress in importance, then re-emerge in adolescence. Children who are five-to-nine years old are in kindergarten and early elementary school where the focus in a classroom is on reading and writing. Children in this age group, who spend most of their days at school, may be allocating more attention to developing these literacy skills and less to AV integration skills. Thus, a larger age-range of participants, where developmental stages before and after this regression period can be represented, would be very beneficial for understanding the impact of AV information on novel word learning across early development.
NH-CISim Outcomes and Cognitive Demand

Although all participant groups demonstrated a significant increase in number of words learned by the end of Block three in the AV task, only children in the NH-CISim group demonstrated improved word learning by the third block of test trials in the AO task. Children in the NH-CISim group had normal hearing, but were presented with spectrally-degraded speech, used to mimic the input transmitted by a CI. The increase in number of words learned during the AO task for this group suggests that these participants may have experienced an increase in cognitive load during this task, as compared to the AV task. For this group of participants, presentation of the degraded auditory input, combined with additional visual speech information, seems to have functioned like a true split attention task (Sweller, 1988; 2011). Due to the increase in cognitive demand, the addition of visual speech cues was not only not useful for improving word learning outcomes but appeared to actually hinder word learning performance.

Differences in NH-CISim and CI Group Outcomes

There were large differences observed within the outcomes of the CI and NH-CISim groups. Children in the NH-CISim group displayed overall better word learning in the AO task, while children in the CI group demonstrated overall better learning outcomes in the AV task. In the current study, the differences observed across the NH-CISim group and CI group may arise from differences in auditory processing. Children in the NH-CISim group had normal hearing, which would suggest typical peripheral and central processing of auditory stimuli. Children in the CI group used at least one CI, which would suggest differences in processing of auditory stimuli, when compared to children with NH. Children in the NH-CISim group may have been able to rely on their typical auditory processing abilities, as well as age-appropriate lexical knowledge, to decipher the degraded auditory input. In other words, children in this group may have been able to utilize top-down processing to more easily decipher and use the degraded input to successfully learn the presented words; thus the addition of visual speech cues was not useful. Children with CIs, whose auditory processing skills differ from children with NH, needed to rely on the degraded auditory input, in combination with the visual speech input, to “fill in the gaps” of the degraded speech to successfully complete this task. Children in the CI group also displayed differences in lexical knowledge as measured through the PPVT-4 (Dunn & Dunn, 2007). Consequently, these children were presumably less able to use established word knowledge, or top-down processing, than their NH peers, to bootstrap word learning in this task. Often, CI-simulated speech is presented to listeners with NH and used to predict or better understand the outcomes of listeners with CIs, across various tasks and assessments. As demonstrated in this study, the outcomes of children with CIs and those with NH, listening to CI-simulated speech, may differ dramatically, and making inferences regarding listeners with CIs from NH-CISim data, should be done with caution.
Variability in Word Learning

Word learning outcomes varied across groups. Children in the NH groups displayed some variability with one (CISim) or two (NoSim) participants achieving less than 50% correct by the end of Block three. Children in the CI group displayed greater variability with half of the participants in this group achieving less than 50% correct by the end of Block three.

Variability Among Participants with CIs

Past studies of vocabulary and language outcomes among children with CIs have demonstrated widely varied outcomes (e.g., Geers et al., 2009; Houston et al., 2001; Lund, 2016; Niparko et al., 2010), and pinpointing factors that may contribute to this variability is particularly challenging (Geers et al., 2016). Prior works which found large discrepancies in language outcomes have subdivided the CI group to encompass ‘good’ and ‘poor’ performance clusters to better characterize factors which contribute to outcomes at an individual level (Eisenberg et al., 2002; Kirk et al., 2007; Lachs et al., 2001). In accordance with prior studies, the CI group in the current study was divided into ‘higher’ and ‘poorer’ performers in an effort to pinpoint and describe factors that may distinguish higher from poor word learners. The two performance clusters of CI participants performed significantly different from one another across the AV and AO tasks in the current study.

Explaining Learning Outcomes

During the AV task, participants had access to multiple AOIs on the presentation screen. The speaker’s face, the story character, and the object to-be-learned were all presented simultaneously across each trial. The second research question in this study was “How do patterns of visual attention during learning relate to word learning outcomes? It was expected that children with CIs and children in the NH-CISim group would spend more time attending to the speaker’s mouth because it provides intersensory redundancy, which should support novel word learning. It was additionally hypothesized that children in the NH-NoSim group would spend more time attending to the object-to-be-learned or the character because children in the NH-NoSim group would not need the additional visual cues, and the character and object may be more perceptually salient.

Visual Attention Patterns

Children in the NH-CISim group demonstrated a shift in their visual attention, away from the speaker’s mouth, and toward the speaker’s eyes as the task progressed, which is inconsistent with the hypothesis. By shifting eye gaze over time to the speaker’s eyes, participants in the NH-CISim group appear to have been actively trying to ignore the additional source of dynamic information, the speaker’s mouth, presumably because
they did not need the support of the visual speech information to understand the degraded input. In addition, shifting attention to the speaker’s eyes “shuts off” input from the stream of additional information (i.e., the speaker’s mouth) thereby reducing cognitive load to optimize performance.

Children in the CI and NH-CISim groups displayed different patterns of visual attention. Children with CIs increased the amount of time they spent attending to the speaker’s mouth across blocks, whereas children in the NH-CISim group decreased the amount of time spent attending to the speaker’s mouth across blocks. This outcome is complementary to the group differences observed in overall learning outcomes across the AV and AO tasks. As previously discussed, children in the NH-CISim group demonstrated better learning during the AO task, possibly due to a lesser cognitive demand than the AV task. The finding that children in this group also shifted their visual attention away from the speaker’s mouth during the AV task, provides further support for the notion that cognitive demand was greater in the AV task for children in the NH-CISim group. Shifting visual attention away from the speaker’s mouth indicates children in this group did not find this additional source of information useful to complete this task, and furthermore may have found it detrimental to performance. In contrast, children in the CI group increased their visual attention to the speaker’s mouth over time, most likely because these participants found it beneficial for better word learning.

Hearing History, Demographic Variables, and Standard Assessment Outcomes

The last research question was, “Which measured factors contribute to novel word learning success across the AV and AO tasks?” For all participants, it was hypothesized that children with bigger receptive vocabularies, better phonological skills, and higher maternal education levels would demonstrate better word learning outcomes. Additionally, for children who use CIs, it was hypothesized that those who received hearing aids earlier in life, were implanted earlier in life, and those who spent more time participating in speech therapy sessions each week, would display better novel word learning outcomes. For all participants, better scores on the standard assessment battery were anticipated to be positively correlated with word learning outcomes. For children with NH, significant correlations were found with subtests of the CTOPP-2 (Wagner et al., 2013) and the TONI-4 (Brown et al., 2010). Demographic factors such as maternal education level and engagement in joint reading were not correlated with novel word learning outcomes.

For children with CIs, multiple subscales of the LEAF (Castellanos et al., 2018) were correlated with outcomes across the Learning tasks, additionally the Blending subtest of the CTOPP-2 (Wagner et al., 2013) was correlated with AO outcomes, and the index score of the TONI-4 (Brown et al., 2010) was correlated with AV task outcomes. No significant correlation with the PPVT-4 (Dunn & Dunn, 2007) was found. The lack of correlation with the measure of receptive vocabulary (PPVT-4) was surprising as multiple prior works have reported relationships among language outcomes, hearing history variables, and receptive vocabulary (e.g., Davidson et al., 2014; Edwards &
Anderson, 2014). There are a couple of possible reasons a correlation was not observed in this study. First, it is possible some of the participants have received a lot of exposure to this assessment, across years of participation in speech therapy. Thus, for this task, the PPVT-4 (Dunn & Dunn, 2007) could have acted as a measure of rote memorization and not receptive vocabulary.

A second possible explanation is the task itself. In the current study, novel word learning was assessed across one or two visits to the laboratory, whereas in a “real life” scenario, novel word learning happens across a greater length of time and with exposure to the novel object-label pair(s) in varying contexts. Success in this task may have been heavily reliant on working memory skills as this task required participants to retain novel word information in short-term memory then access that information during the Test phase of the task. If this task had examined word learning over a longer period of time, where participants needed to move the new word information into long-term memory and integrate it with existing word knowledge, perhaps a correlation would have emerged. An objective assessment of working memory would be needed to further investigate the relationship among these variables.

Performance Differences Among Participants with CIs

Of the hearing history variables investigated, only age of amplification was significantly related with AO learning outcomes when the CI group was intact. When participants with CIs were subdivided into performance groups, age of amplification, age of implantation, and phonological processing skills emerged as predictors for performance.

Word learning differences across the performance subgroups (high and poor) were correlated with differences in looking patterns. Specifically, the proportion of the Block three Test trial spent looking to the speaker’s mouth was correlated with learning outcomes by the end of Block three where ‘high’ performers spent significantly longer looking to the speaker’s mouth than ‘poor’ performers, as shown in Figure 4-8. Participants in the ‘poor’ performers group appeared to also spend more time attending to the speaker’s mouth as the Test phase continued from Block one to Block three, but the proportion of time spent looking to the speaker’s mouth grew at a slower rate than children in the ‘high’ performers group. This trend, as well as the differences noted in Block three, suggest that the higher performers may have developed a learning strategy of attending to the speaker’s mouth and were able to better use these visual speech cues than poorer performers. Perhaps children who were poorer performers were either unable to utilize such a strategy or developed a successful word learning strategy at a slower pace. It may be that with additional time and exposure to the object-label pairs, children in the ‘poor’ performers group would continue to increase looking time to the speaker’s mouth and display learning outcomes more similar to children in the ‘high’ performers group.

Outcomes across the ‘high’ and ‘poor’ performance groups could be attributed to differences in age of amplification, age of implantation, and phonological processing
skills. Children noted as ‘high’ performers received amplification (via hearing aid) and their implant(s) earlier in life than children who were noted as ‘poor’ performers. Differences also stemmed from outcomes on the Blending subtest of the CTOPP-2 (Wagner et al., 2013). The Blending subtest required participants to maintain phonemes presented over time, in short-term working memory, then piece the phonemes together to recall the entire word. Therefore, this task not only required accurate perception of individual phonemes without visual speech information, but also robust working memory skills.

**Differences Across Word Learning and Sentence Recognition Tasks**

Although significant differences in word learning across tasks were not found, children with CIs displayed significantly better sentence recognition scores when the sentences were presented in an AV format, as compared to outcomes across the AO sentence recognition task (MLST-C; Kirk et al., 2012). Additionally, performance on the MLST-C was not significantly correlated with novel word learning outcomes across either task (AV or AO). The lack of significantly different outcomes across task format (AV and AO) in the learning assessment suggests that the task demand differences across the sentence recognition and novel word learning tasks may alter the perceptual benefit of visual speech cues. Although participants with CIs did demonstrate better word learning in the AV word learning task by the end of Block three, as they did in the AV sentence recognition task (MLST-C), the amount AV gain achieved was much greater during the sentence recognition task than in the novel word learning task.

**Clinical Implications**

All participants in the CI group used spoken language and participated in speech therapy that focused on the use of spoken English. Clinically, it has been assumed that allowing children with CIs to use visual speech cues would inhibit their ability to use auditory input from their device(s). This unisensory perspective has guided some clinical practices, including blocking the speaker’s mouth when conversing with a child with a CI to prevent the child from using visual speech cues from the speaker. Recently, with evidence that children with CIs can utilize, and benefit from visual speech information, a multisensory approach has emerged (see McDaniel et al., 2018 for a review). In a multisensory approach, utilization of AV speech cues is encouraged. Although it is important to acknowledge that blocking visual speech cues for verification of device functioning is important, during learning tasks, it may be most beneficial to allow children with CIs access to visual speech information. Access to AV speech information may not only support accurate speech perception, but additionally encourage an adaptive strategy development for word learning.

The differences in visual attention across the higher and poorer CI performers, observed in this task, suggest differences in word learning strategies. Across trials, higher performers not only spent a greater proportion of time attending to the speaker’s mouth,
but allocated more of their attention to the speaker’s mouth sooner than poorer performers. Although poorer performers appeared to shift their attention later in the task, participants in this subgroup continued to attend to the speaker’s mouth over time, which would suggest these participants were able to use this information but perhaps needed additional time and exposure to achieve learning outcomes akin to participants in the higher performers subgroup.

**Limitations**

Although recent changes to implantation criteria and more widespread newborn hearing screening programs have led to a larger number of children receiving CIs, ‘children with CIs’ is still a low incidence population. Like many studies of children with CIs, the current study had a small sample size of 13 children (12 included in analyses). A larger sample size may have introduced more diversity across the CI population to possibly provide more information of factors that contribute to learning across this group and allow of the analysis of factors such as communication mode as each relates to novel word learning outcomes. Additionally, the underpowered sample may also have prohibited an observation of differences in learning outcomes across tasks, specifically in Block three.

Another limitation worth noting is the lack of ecological validity of studying novel word learning in a controlled and quiet laboratory setting. Children spend most of their day in classrooms, which are notoriously noisy (Crandell & Smaldino, 2000), and children are often further than one meter from the person they are listening to. Further investigations are necessary to understand the influence of listener-to-speaker distance and background noise on novel word learning, and how these factors may alter the benefit and utilization of visual speech information.

**Conclusions**

Children with CIs displayed a wide range of word learning outcomes but did not demonstrate a significant difference in learning across AV and AO tasks, suggesting that the introduction of access to visual speech cues will not always improve novel word learning performance. Within the CI group, ‘high’ and ‘poor’ performers emerged and performed significantly different from one another on the learning tasks and additionally displayed different patterns of visualizations during the Learning phase. Association with the ‘high’ and ‘poor’ performance groups could be attributed to age of amplification, age of implantation, and phonological processing skills as measured through the Blending subtest of the CTOPP-2.

Children with NH, specifically those in the NH-CISim group, displayed different patterns of learning and visual attention than anticipated. These differences may be due to a shift in the cognitive load across the Learning tasks. CI-simulated speech may be a valuable tool for investigating cognitive load during listening in future work. Future work
should consider a more ecologically valid task paradigm, such as the introduction of background noise and different speaker-listener configurations in physical space. Additional measures of executive functioning apart from the parental report may provide more in-depth information into central skills, like working memory, which presumably contribute to learning outcomes, as observed with the Blending subtest of the CTOPP-2 in the current study.


Brown, L., Sherbenou, R.J., Johnsen, S.K. (2010). Test of Nonverbal Intelligence (4th ed.). Austin, TX: PRO-ED.


VITA

Kristen Elizabeth Thompson Thornton was born in 1991 in Beckley, West Virginia. After finishing high school at Independence High School she started school at Concord University in Athens, West Virginia where she received a Bachelor of Arts in Psychology and Sociology with a minor in Biology. In 2012 she started school at the University of Tennessee in Knoxville, Tennessee. Kristen earned her Master of Arts degree in Experimental Psychology in 2014. The following semester Kristen started the PhD program in Speech and Hearing Science at the University of Tennessee Health Science Center in Knoxville, Tennessee. Kristen is currently an Assistant Professor at Gallaudet University’s Hearing, Speech, and Language Sciences program in Washington, DC.