8-2018

Changes in Audiovisual Word Perception During Mid-Childhood: An ERP Study

Elizabeth Ancel
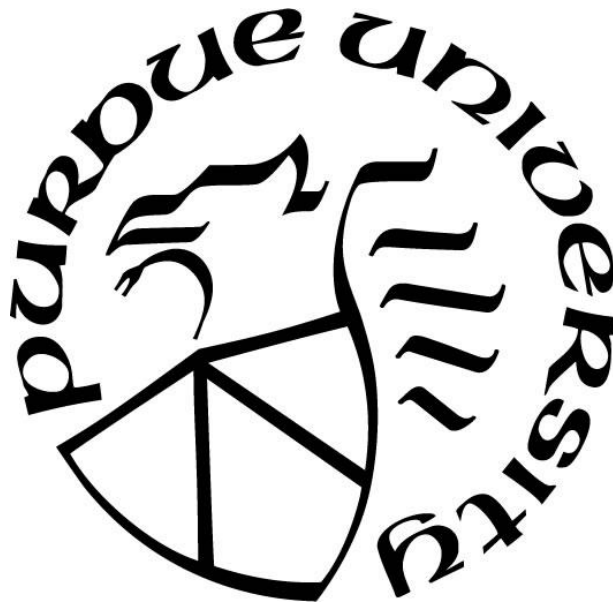*Purdue University*

CHANGES IN AUDIOVISUAL WORD PERCEPTION DURING MID-CHILDHOOD: AN ERP STUDY

by

Elizabeth Ancel

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Speech, Language, & Hearing Sciences

West Lafayette, Indiana

August 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Natalya Kaganovich, Chair

      Department of Speech, Language, & Hearing Sciences

Dr. Brandon Keehn

      Department of Speech, Language, & Hearing Sciences

Dr. Laurence Leonard

      Department of Speech, Language, & Hearing Sciences

Approved by:

      Dr. Keith Kluender

        Head of the Graduate Program

Dedicated to the Purdue Speech-Language Pathology Class of 2018, my family, and my mentors, Dr. Natalya Kaganovich and Jen Schumaker.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Ancel, Elizabeth, E. MS
Institution: Purdue University
Degree Received: August 2018
Title: Changes in Audiovisual Word Perception During Mid-Childhood: An ERP Study
Major Professor: Kaganovich, Natalya

Throughout school-age years, speech perception is an important skill that often relies on the child's ability to combine both auditory and visual information from the speaker. In order to better understand the development of multisensory speech perception during mid-childhood, we analyzed audiovisual word perception in three groups of participants: 8-9-year olds, 11-12-year olds, and adults. Participants matched visually-perceived articulatory movements with corresponding auditory words. In "congruent" trials, the auditory word matched the subsequently presented silent visual articulation. In "incongruent" trials, the words presented differed on the initial phoneme. From this task, we evaluated specific neural components —the N400 and the Late Positive Complex (LPC) — which index the phoneme and whole word level of audiovisual processing, respectively. The results of this experiment were then related to a real-life behavioral speech perception skill, namely, listening to speech-in-noise. Our results suggest that while the LPC becomes adultlike by the age of 11 or 12, the N400 is not fully matured until later in development. In addition, the relation of the LPC to listening to speech-in-noise is stronger earlier in childhood while the relation of the N400 is stronger during later school years and adulthood. Overall, we show that audiovisual processes related to the whole-word level mature earlier in childhood than processes related to the phonological level.

CHAPTER 1. INTRODUCTION

Speech perception in children and adults is audiovisual in nature. Well documented phenomena such as the McGurk Illusion and the Ventriloquist Illusion are clear demonstrations of how much listeners rely on visual information when listening to speech. For the McGurk Illusion, perception of the same auditory signal is changed with the addition of different articulatory movements. As an example, a person may perceive the sound "da" if they hear the sound "ba" while at the same time seeing someone articulating "ga." In the Ventriloquism Illusion, the source of a speech signal is affected based on the location of visual speech cues, creating a strong sensation that the voice's source is the puppet and not the puppeteer. Both of these examples suggest that listeners partially base interpretations of auditory input on the associated visual information they receive.

There are numerous examples of the importance of being able to integrate visual information along with the auditory signal. Across populations, listeners perceive speech more accurately when speech is audiovisual, as opposed to only auditory (Sumby & Pollack, 1954). Cochlear implant (CI) users present a particularly insightful case because of their unique perceptual experience across the lifespan. Specifically, CI users are generally exposed to a lower quantity of auditory information, and that information is lacking full frequency resolution. As a result of this experience, adult CI users show better speechreading ability than typically hearing adults, which supports the idea that they have improved visual speech perception to compensate for decreased auditory information (Pimperton, Ralph-Lewis, & MacSweeney, 2017). More importantly, there is also evidence that CI users have better audiovisual speech perception abilities as a result of more reliance on visual speech cues. In one study by Rouger and

colleagues (2007), CI users and typically hearing subjects were presented similar auditory words, meaning that the words presented to the typically hearing subjects were degraded to simulate listening with a CI. In this context, it was found that CI users have more accurate word recognition performance with the addition of visual information compared to typically hearing subjects. Furthermore, a study by Kirk and colleagues (2002) demonstrated that adult CI users were able to improve their speech perception at the word level significantly more than typical adults when additional visual speech cues were from a single talker. These results show that experience with visual speech cues improves perception of the auditory speech signal.

In addition to CI users, older adults have been shown to have changes in audiovisual speech processing as compared with typical, younger adults. For example, in a study by Thompson and Malloy (2004) that tracked the ability to detect small dots on a person's eyes and mouth while listening to them give a speech, older adults were relatively better at detecting dots around the speaker's mouth as compared to younger adults, suggesting that older adults are continuously looking more at a speaker's mouth while comprehending audiovisual speech. In a related finding, it has also been shown that both older adults and CI users have an increased susceptibility to the McGurk illusion (Rouger, Fraysse, Deguine, & Barone, 2008; Schorr, Fox, Van Wassenhove, & Knudsen, 2005; Sekiyama, Soshi, & Sakamoto, 2014), further supporting the increased usage of visual speech cues in these groups compared to typical adults. A study by Stevenson et. al. (2015) further explored multisensory speech perception by evaluating younger and older adults on their word-level and syllable-level audiovisual speech perception abilities at different signal-to-noise ratios (SNRs). At the word level, the older adults differed from younger adults in that they did not have more gain from seeing the talker's face at low SNRs, showing degradation in audiovisual speech processing in that context. However, the older adults had

similar gain as typical adults when the same conditions were applied to the syllable level. Overall, we see that audiovisual speech perception can be affected by a wide range of factors including age, noise level, and experience. Importantly, we see that audiovisual speech perception is also affected by whether the person is processing at the word or at the syllable level.

The reliance on visual speech cues is not only found in groups like older adults and CI users. Typical, young adults also tend to focus more on a speaker's mouth under challenging listening conditions. A study by Vatikiotis-Bateson and colleagues (1998) found that when listeners were presented audiovisual monologues with masking auditory noise in the background, they increased the amount of time spent looking at the mouth. Other studies suggest that seeing the talker's face is also helpful when listening to speech spoken in a second language. In one study by Navarra and Soto-Faraco (2007), participants were able to accurately distinguish between two phonemes in a second language only when the presentation was audiovisual, as opposed to unimodal. In sum, the ability to combine visual cues with auditory information is beneficial to typically developing young adults.

While adults are adept at comprehending speech and using visual information in tandem, children and adolescents are continually developing their abilities in these areas. For example, a study by Elliot (1979) of children at ages 9, 11, 13, 15, and 17 years examined the participants' ability to understand spoken words embedded in babble at different SNRs. In that experiment, the 13-year-olds performed significantly worse than the 15- and 17-year-olds, and 9-year-olds performed significantly worse than 11-year-olds. Only the 17-year-olds performed similarly to adults. These results showed that while children's hearing was fully intact, they still tended to struggle significantly more when babble was present. Similarly, a study by Papso and Blood

(1989) showed that 4- to 6-year-old children performed worse than adults when identifying words in multitalker babble and in pink noise. More recently, Ross and colleagues (2011) evaluated the ability of 5- to 14-year-old children and of adults to understand speech-in-noise with and without the presence of the talker's face. In agreement with earlier studies, these authors reported that younger children benefited significantly less than adults from being able to see the talker's face. However, the 12- to 14-year-old group had similar gain from the added visual information as compared with adults. In sum, earlier studies suggest that audiovisual speech perception is not fully developed until later in adolescence. Although we know that audiovisual skills change significantly during school years, we do not yet know which neural and cognitive processes underlie this development or when this development takes place.

During school years, the brain experiences an increase in synaptic growth before the onset of puberty and another period of maturation during adolescence (Jetha & Segalowitz, 2012). In addition, it is known that the rate of maturation is distinct for different brain areas (Murray, Lewkowicz, Amedi, & Wallace, 2016). Because of the variation in brain development over time, it is important to understand the maturation of specific task-related skills throughout childhood. This study aims to track the development of specific neural processes related to the perception of visual speech gestures in school-age children.

Previous studies in our lab focused on two event-related potential (ERP) components related to audiovisual speech processing- the N400 and the Late Positive Complex (LPC). The N400 component is a negative waveform occurring at approximately 400 ms post-stimulus onset. While the N400 has been studied most extensively as a marker of semantic processing, it is also sensitive to phonological incongruency. For example, in one study by Praamstra and Stegeman (1993), participants were presented with pairs of words and non-words and then directed to

decide whether the pairs rhymed. Non-rhyming pairs produced an N400 after presentation of the second word, indicating that participants recognized a mismatch between the phonological components of the words. This effect was present regardless of whether the words were produced by different speakers and regardless of whether the words were real or made-up. This provides evidence that the N400 amplitude is sensitive to mismatches and novel information on a phonological level.

The LPC component is a positive waveform occurring at approximately 600 ms or later following the stimulus onset. While the N400 is reactive to new stimuli, the LPC has been shown to be modulated by recognition memory. The LPC is sensitive to repeated stimuli presented in a variety of modalities. A study from Rugg and Nagy (1987) further explored the LPC by investigating its responsiveness to "legal" nonwords that have acceptable English orthography (e.g. "FRIME") and "illegal" nonwords that have unacceptable English orthography (e.g. "SKHRA"). They found that responses to the repetition of legal nonwords were larger during the 402-600 ms range, regardless of whether the task was counting real words or counting words containing an "@" symbol. However, no LPC was observed to the repetition of illegal nonwords. Because legal nonwords inherently have more lexical relevance than illegal nonwords, this suggests that the LPC is sensitive to the repetition of items with stronger lexical representations. In other words, the LPC amplitude is modulated by repetition at the whole-word level.

In earlier studies in our lab, we examined the N400 and the LPC in response to audiovisual word matching in children and adults. Participants were asked to determine whether spoken words matched a subsequently presented silent articulation of a word. During "congruent" trials, the spoken word and the video of the articulation depicted the same word, while during "incongruent" trials, the words differed on the first phoneme. This experiment was

combined with an ERP recording in order to analyze each participant's neural responses to the video of the silent articulation. These neural responses were then evaluated with regard to behavioral data from an SIN task administered in a separate session.

In these earlier studies, the N400 and LPC have shown distinct responses to the paradigm described above. The N400 has been found to be larger to "incongruent" trials in which the word presented auditorily and the silent articulation were a mismatch (Kaganovich, Schumaker, & Rowland, 2016b). In light of the findings from Praamstra and Stegeman (1993), the participants were thought to have detected incongruencies between the stimuli on a phonological level, as opposed to a lexical level. Combined with the fact that the N400 occurs shortly after the initial articulatory gesture of the silent word, instead of after the completion of the word, it can be inferred that it is responding to phonological information present at the beginning of the word. Unlike most rhyming studies conducted before, our paradigm is audiovisual, meaning that it relies on the correspondences in phonological information conveyed by sound and by sight. It requires the participant to integrate phonological information obtained first from an exclusively auditory modality and then from a purely visual representation. In this way, we are able to evaluate each participant's *multisensory* sensitivity to phonological information through the modulation of the N400 amplitude.

As mentioned above, the LPC is usually enhanced by a repeated presentation of a word (Rugg & Curran, 2007). In a similar way, "congruent" trials of our studies contained a repeated presentation of the same word. However, the original presentation was an auditory word, and the repetition of that stimulus was presented with an articulatory gesture. Therefore, although the physical aspects of the first and second presentations are completely different, they share a

higher-level lexical representation. Accordingly, visually-perceived articulations that matched previously heard auditory words elicited larger LPCs compared to articulations that did not.

Together, the N400 and the LPC can be used to track the developmental trajectory of audiovisual word perception at two different levels of linguistic analysis. This study asks two main questions: *How does audiovisual integration mature behaviorally and neurophysiologically throughout childhood?* and *How does the maturation of neural processes involved in audiovisual integration relate to real-life skills such as speech-in-noise (SIN) perception?*

These questions are important for exploring the typical maturation during school age years. It is not known whether the N400 and the LPC have similar or distinct developmental trajectories. In addition, it is also not known how these components relate to real-life skills such as SIN perception in children. Children are consistently exposed to environments, such as noisy classrooms, that present difficult listening conditions. In order to fully understand children's speech perception abilities in difficult listening situations at varying ages, it is important to consider the development of neural processes and relevant real-life skills.

This understanding is also important in the context of children who are not typically developing. For children who may have deficits in audiovisual integration or speech perception, it is important to have a well-rounded perception of what those skills look like in typically developing children. Further, although some children may be functionally within the normal range on audiovisual tasks, they may still show deficits or abnormal processing patterns on a neural level (Annaz, Karmiloff-Smith, & Thomas, 2007). Lastly, strengths and weaknesses in a child may change over time. Gaining insight into the full course of development of neural processes specific to audiovisual speech perception can help us recognize abnormalities at a wider range of ages throughout childhood.

CHAPTER 2. METHODS

2.1 Participants

Participants were divided into three groups, each with 14 members. The younger group of children consisted of 8-9 year-olds (8 females, mean age 9;0, range: 8;0-9;11). The older group of children consisted of 11-12 year-olds (5 females, mean age 11;9, range 11;1-12;9). Lastly, a group of adults (8 female, mean age 23, range 18-37) was included for comparison. All participants over 18 gave their written consent to participate in the experiment, and child participants gave their written assent. Parental consent was also obtained for child participants. The study was approved by the Institutional Review Board of Purdue University and all study procedures conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki, 1964).

All participants were assessed to identify any atypical physical or neurological characteristics. We evaluated verbal working memory with the non-word repetition test (Dollaghan & Campbell, 1998) and the Number Memory Forward and Number Memory Reversed sub-tests of the Test of Auditory Processing Skills – 3rd edition (TAPS-3, Martin & Brownel, 2005). All participants were administered the Test of Nonverbal Intelligence - 4th edition (TONI-4, Brown, Sherbenou, & Johnsen, 2010) to rule out intellectual disability. Handedness was assessed with an augmented version of the Edinburgh Handedness Questionnaire (Cohen, 2008; Oldfield, 1971). In the 11-12 group, one participant was left-handed and one was ambidextrous. In the adult group, two participants were ambidextrous. All other participants were right-handed. None of the participants had any history of atypical language development, ADHD, or reading difficulties. All were free of neurological disorders

(e.g., seizures), passed a hearing screening at a level of 20 dB HL at 500, 1000, 2000, 3000, and 4000 Hz, and reported to have normal or corrected-to-normal vision.

The child participants underwent additional testing to identify any indications of atypical development. We administered 4 subtests of the Clinical Evaluation of Language Fundamentals (CELF-4, Semel, Wiig, & Secord, 2003) to each child in order to assess their current language ability - the Concepts and Following Directions (C&FD), Recalling Sentences (RS), Formulated Sentences (FS), Word Structure (WS, 8 year olds only), and Word Classes (WC-2, 9-12 year olds only). Taken together, these subtests yielded the Core Language Score (CLS) and the Expressive Language Index (ELI), which reflect general linguistic aptitude and language production. In addition, all children were evaluated with the Childhood Autism Rating Scale - 2nd edition (CARS-2, Schopler, Van Bourgondien, Wellman, & Love, 2010) to rule out the presence of autism spectrum disorders. The level of risk for developing ADHD was evaluated with the help of the short version of the Parent Rating Scale of the Conners' Rating Scales – Revised (Conners, 1997). The level of mothers' education was measured as an indicator of children's socio-economic status (SES).

Table 1 contains group means, standard deviations, and the outcomes of group comparisons for measures of non-verbal intelligence (based on TONI-4) and verbal working memory (based on the Number Memory Forward and Number Memory Reversed subtests of the TAPS-3), administered to all groups. None of the group comparisons were significant.

Table 1: Group means for age, verbal working memory (TAPS-3), and non-verbal intelligence (TONI-4).

	8-9 year olds	11-12 year olds	Adults	*F (df1,df2)*	*p*
Age (years; months)	9;0 (0.66)	11;9 (0.62)	22;10 (4.57)	104.16 (2, 41)	<.01
TAPS-3 number memory					
Forward	9.79 (2.49)	9.57 (1.99)	11.14 (2.57)	1.82 (2, 41)	.18
Reversed	11.86 (2.85)	10.07 (2.97)	11.86 (3.06)	1.69 (2, 41)	.20
TONI-4	106.64 (9.83)	110.07 (11.16)	103.36 (6.48)	1.80 (2, 41)	.18

Note. Numbers for TONI-4 and TAPS-3 represent standard scores, with the mean of 100 and a standard deviation of 15 for TONI-4 and a mean of 10 and standard deviation of 3 for TAPS. Numbers in parenthesis are standard errors of the mean. *F* and *p* values reflect a group comparison.

Table 2 contains group means and standard deviations for measures of SES (maternal and paternal education level), linguistic ability (CELF-4 Core Language Score and Expressive Language Index), presence of Autism Spectrum Disorder (ASD) (CARS-2), and presence of ADHD symptoms (Connors' Rating Scales) administered to both child groups. The groups did not differ in their mothers' education levels, fathers' education levels, presence of ASD, or presence of ADHD symptoms. Data for paternal level of education was not available for three subjects (one subject in the 11-12-year old group and two subjects in the 8-9-year old group). They also did not differ in linguistic ability. Group comparisons of the Word Structure and Word Classes subtests were not conducted since the groups contained different ages, and therefore took different subtests.

Table 2: Child group means for parents' education levels (SES), linguistic ability (CELF-4), presence of autism (CARS-2), and presence of ADHD symptoms (Connors' Rating Scales).

	8-9 year olds	**11-12 year olds**	*F (df1,df2)*	*p*
Mother's Education (years)	14.86 (2.32)	15.43 (1.83)	0.53 (1, 27)	.48
Father's Education (years)	15.83 (3.46)	17.23 (3.60)	0.97 (1,24)	.33
CELF-4				
CF&D	11.07 (1.64)	11.50 (1.45)	0.54 (1, 27)	.47
RS	11.07 (2.09)	10.43 (1.99)	0.69 (1, 27)	.41
FS	12.71 (1.77)	12.14 (2.03)	0.63 (1,27)	.44
CLS	111.07 (8.86)	110.21 (9.07)	0.06 (1,27)	.80
ELI	111.93 (9.71)	108.21 (11.18)	0.88 (1,27)	.36
CARS-2	15.21 (0.58)	15.00 (0.00)	1.92 (1, 27)	.18
Connors' ADHD Index	50.79 (5.77)	47.71 (6.29)	1.81 (1, 27)	.19

Note. Data for paternal level of education was not available for three subjects (one subject in the 11-12-year old group and two subjects in the 8-9-year old group). Numbers for Connors' ADHD Index, CLS, ELI, and CARS-2 represent standard scores. Numbers in parenthesis are standard errors of the mean. *F* and *p* values reflect a group comparison. CF&D Concepts and Following Directions, RS Recalling Sentences, FS Formulated Sentences, CLS Core Language Score, ELI Expressive Language Index.

2.2 Experiment 1- Audiovisual Matching Task

2.2.1 Stimuli

Stimuli for experiment 1 consisted of auditory words, silent videos of their articulations, and pictures matching words' meanings. We used 96 words from the MacArthur Bates Communicative Developmental Inventories (Words and Sentences, Fenson et al., 2007) as stimuli. All words contained 1-2 morphemes and were 1 to 2 syllables in length with two exceptions – "elephant" and "teddy bear." Words contained between 1 phoneme (e.g. eye) and 8 phonemes (e.g. sandbox), with diphthongs counted as 1 phoneme. Words were produced by a female speaker and recorded with a Marantz digital recorder (model PMD661) and an external microphone (Shure Beta 87) at a sampling rate of 44,100 Hz. Sound files were edited in the Praat software (Boersma & Weenink, 2011) so that the onset and offset of sound were preceded by 50 ms of silence. Final sound files were root-mean-square normalized to 70 dB.

Videos showed a female talker dressed in a piglet costume articulating one word at a time. The costume made it easier to turn the paradigm into a game and to maintain children's attention. The actor's mouth area was left free of makeup except for bright lipstick that did not obscure natural muscle movements of the lower face during articulation. The videos' frame per second rate was 29.97. The audio track of the video recording was removed in Adobe Premier Pro CS5 (Adobe Systems Incorporated, USA). Articulation portions of videos ranged from 1133 ms (for "car") to 1700 ms (for "sandbox").

Each of the words was matched with a color picture from the Peabody Picture Vocabulary Test (pictures were used with the publisher's permission, Dunn & Dunn, 2007) that exemplified the word's meaning (for example, a picture of toys was matched with the word "toys"). Pictures served as fixation points to better maintain children's attention on the computer monitor and minimize eye movements.

2.2.2 Experimental Design

The experimental design was identical to that described in two earlier studies from our laboratory (Kaganovich, Schumaker, & Rowland, 2016a; Kaganovich et al., 2016b). Each trial consisted of the following events (see Figure 1). Participants saw a color picture of a common object/person (e.g., toys, mailman, etc.). While the image was on the screen, participants heard the object named (e.g., they heard a female speaker pronounce the word "toys" or "mailman," etc.). A blank screen followed for 1000 ms. Next, a video of a female talker was presented. It consisted of a static image of the talker's face taken from the first frame of the video (1000 ms), followed by a silent articulation of a word, followed by the static image of the talker's face taken from the last frame of the video (1000 ms). In half of all trials, the talker's articulation matched the previously heard word (*congruent* trials; for example, participants saw the talker articulate

"toys" after hearing the word "toys"), while in another half, the talker's articulation clearly mismatched the previously heard word (*incongruent* trials; for example, participants saw the talker say "bus" after hearing the word "toys"). Next, the appearance of the screen with "Same?" written across it signaled the start of the response window. It lasted 2000 ms, during which participants had to determine whether the silently articulated word was the same as the word they heard at the beginning of the trial. Trials were separated by a temporal period randomly varying between 1000 and 1500 ms. Responses were collected via a response pad (RB-530, Cedrus Corporation), with the participants instructed to press a green button for perceived "congruent" trials and a red button for "incongruent" trials. The response hand to each trial type was counterbalanced across participants. Stimulus presentation and response recording was controlled by the Presentation program (www.neurobs.com).
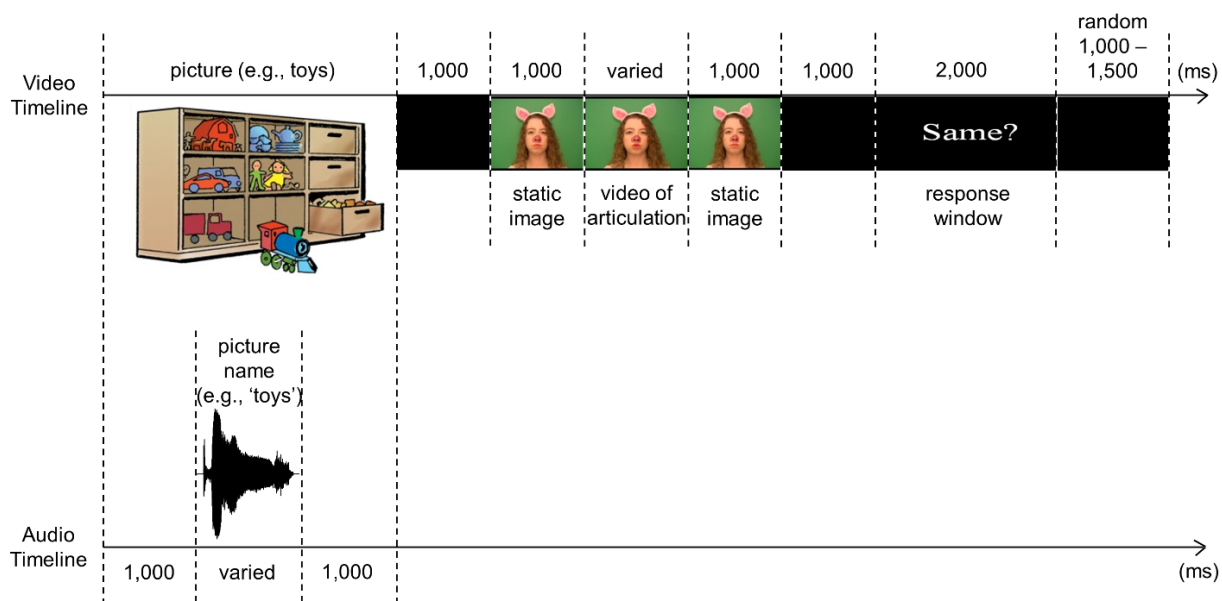
Figure 1: Schematic Representation of an Audiovisual Matching Task Trial. Note that separate timelines are shown for the video and audio tracks. The video of articulation was congruent in half of all trials (e.g. participants saw the piglet silently articulate "toys" after hearing "toys" at the start of the trial) and incongruent in the other half of trials (e.g. participants saw the piglet silently articulate "bus" after hearing "toys" at the start of the trial). The onset of articulation was used as time 0 for ERP averaging.

Each participant completed 96 trials (48 congruent and 48 incongruent). For incongruent trials, 48 pairs of auditory and silently articulated words were created such that their visual articulation differed significantly during the word onset. In most cases (35 out of 48 pairs), this was achieved by pairing words in which the first consonants differed visibly in the place of articulation (e.g., belt vs. truck). In 6 pairs, the first vowels of the words differed in the shape and the degree of mouth opening (e.g., donkey vs. candy). In the remaining 7 pairs, the first sounds were a labial consonant in one word (i.e., required a mouth closure (e.g., pumpkin)) and a vowel (i.e., required a mouth opening (e.g., airplane)) in another word. Heard and articulated words in incongruent pairs had no obvious semantic relationship. Two lists containing 48 congruent and 48 incongruent heard versus articulated word presentations were created such that articulations that were congruent in list A were incongruent in list B. As a result, across all participants, we collected responses to the same articulations, which were *perceived* as either congruent or incongruent. Such counterbalancing also allowed for the control of word frequency, length, and complexity across congruent and incongruent trials. Lastly, 10 different versions of list A and 10 different versions of list B were created by randomizing the order of 96 trials. Each participant completed only one version of one list (e.g., participant 1 did list A version 1; participant 2 did list B version 1; participant 3 did list A version 2, participant 4 did list B version 2, etc.) Version 1 of lists A and B is shown in the Appendix. This task was combined with ERP recordings (see below).

2.2.3 ERP Recordings and Data Analysis

2.2.3.1 General Procedure

During the audiovisual matching task, the electroencephalographic (EEG) data were recorded from the scalp at a sampling rate of 512 Hz using 32 active Ag-AgCl electrodes secured

in an elastic cap (Electro-Cap International Inc., USA). Electrodes were positioned over homologous locations across the two hemispheres according to the criteria of the International 10-10 system (American Electroencephalographic Society, 1994). The specific locations were as follows: midline sites Fz, Cz, Pz, and Oz; mid-lateral sites FP1/FP2, AF3/AF4, F3/F4, FC1/FC2, C3/C4, CP1/CP2, P3/P4, PO3/PO4, and O1/O2; and lateral sites F7/F8, FC5/FC6, T7/T8, CP5/CP6, and P7/P8; and left and right mastoids. EEG recordings were made with the Active-Two System (BioSemi Instrumentation, Netherlands), in which the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode replace the traditional "ground" electrode (Metting van Rijn, Kuiper, Dankers, & Grimbergen, 1996). Data were referenced offline to the average of the left and right mastoids. The Active-Two System allows EEG recording with high impedances by amplifying the signal directly at the electrode (BioSemi, 2013; Metting van Rijn et al., 1996). In order to monitor for eye movement, additional electrodes were placed over the right and left outer canthi (horizontal eye movement) and below the left eye (vertical eye movement). Prior to data analysis, EEG recordings were filtered between 0.1 and 30 Hz. Individual EEG records were visually inspected to exclude trials containing excessive muscular and other non-ocular artifacts. Ocular artifacts were corrected by applying a spatial filter (EMSE Data Editor, Source Signal Imaging Inc., USA)(Pflieger, 2001). ERPs were epoched starting at 200 ms pre-stimulus and ending at 1800 ms post-stimulus onset. The 200 ms prior to the stimulus onset served as a baseline.

2.2.3.2 ERP Components Measured

The onset of articulation elicited clear N400 and LPC. These components' mean amplitudes were measured over the following windows: 430-750 ms for N400 and 930-1540 ms for LPC. N400 and LPC were initially measured over the CP, P, PO, and O sites. After analysis

of each site, it was determined that the O site differed significantly from the other sites for the

N400 waveform. Because of this difference and because the O site is not always used for

visualization of the N400 component, it was excluded from later analyses. The sites selected and

windows for the N400 and LPC components were based on previous studies of adults and

school-age children tested with the same paradigm (Kaganovich et al., 2016a, 2016b). To

encompass components at different ages, windows were lengthened based on visual inspection of

the grand average waveforms.

2.3 Experiment 2 – Word Identification Tasks

2.3.1 Experimental Design Overview

In the second experiment, participants were presented auditory words and videos of

corresponding articulations from the list of 96 words used in the audiovisual matching task. This

experiment consisted of two different tasks— a SIN task and a lip-reading task. In the SIN task,

words were presented in two different conditions—auditory only (A) and audiovisual (AV). In

the lip-reading task, words were presented in a visual only (V) manner. Participants were asked

to verbally identify each word to the best of their ability.

2.3.1.1 Speech-In-Noise Task.

A schematic representation of the SIN task is shown in Figure 2. In the AV condition,

participants saw videos of a talker producing each of the 96 words. Each video was preceded and

followed by a static image of a talker with a closed mouth, which lasted for 1,000 ms. In the A

condition, the same static images of the talker were present; however, the video portion was

replaced with an image of the talker with her mouth open (see Figure 2). The appearance of the

open-mouth picture in the A condition cued participants to the onset of the auditory word,

without providing any visual cues to its identity. Previous research shows that visual cues that

reliably predict the onset of the auditory signal significantly improve the latter's detection threshold (ten Oever, Schroeder, Poeppel, van Atteveldt, & Zion-Golumbic, 2014). The inclusion of the cue to the target word onset in the A condition aimed to make the attentional demands of the A and AV conditions more similar. Word presentations in both conditions were separated by 3 seconds, during which participants provided their verbal response about what they had heard. When unsure, participants were encouraged to give their best guess or to say "I don't know." These two conditions contained all 96 words each and were administered on two separate days. The order of A and AV conditions was counterbalanced across participants, but each participant completed both.

Recordings of the words were embedded in a two-talker babble masker. The masker consisted of two female voices reading popular children's stories. One sample was 3 minutes and 8 seconds long (by talker 1), and the other was 3 minutes and 28 seconds long (by talker 2). Both samples were manually edited in Praat to remove silent pauses greater than 300 ms and then repeated without discontinuity. The streams from the two talkers were root-mean-square normalized to 75 dB, mixed, and digitized using a resolution of 32 bits and a sampling rate of 24.414 kHz. Because 96 words were root-mean-square normalized to 70 dB, the final stimuli had a -5 dB signal-to-noise ratio. The babble masker started 3 seconds prior to the first trial and was presented continuously until the end of the experiment.
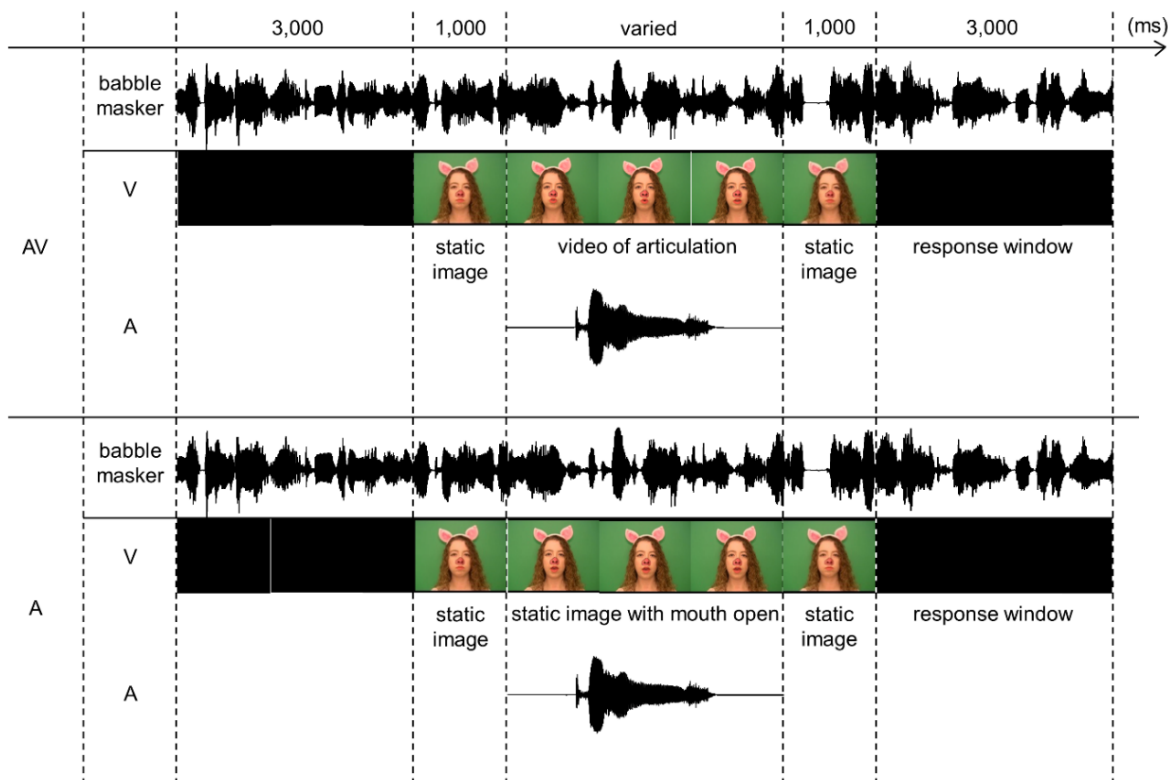
Figure 2: Schematic Representation of a Speech-In-Noise (SIN) Trial. The SIN task had two conditions- the audiovisual (AV, top panel) and the auditory only (A, bottom panel). Note that separate timelines are shown for the video and audio tracks in each condition. In the AV condition, participants saw a video of the piglet articulating target words, while in the A condition the video was replaced with a static image of the piglet's face with her mouth open. The appearance of the open mouth picture in the A condition cued participants to the fact that the onset of the auditory word was imminent, but provided no visual cues to its identity.

2.3.1.2 Lip-reading Task

The lip-reading task was used to determine how many of the silent articulations could be recognized by our participants on incongruent trials in the audiovisual matching task and to evaluate their lip-reading abilities (which are often thought to contribute to SIN perception). We selected 20 silent articulations from the list of 96 used and asked each participant (in a separate session from the audiovisual matching task) to provide their best guess as to what word they thought the speaker was producing. The list of 20 words used for this task is shown in Table 3. In order to select words that reflected the diversity of lexical items used for the main task, this set of

words included both one- and two-syllable words and contained items that started with either a labial (closed mouth) or an alveolar (open mouth) sound. No cues to the words' identity were provided. This task is referred to henceforth as the lip-reading task. Because in many cases multiple auditory words may map onto similar observable articulatory movements, participants were given credit not only for identifying the word that was in fact produced by the talker but also for reporting words that shared the same articulation with the target word. For example, the words "Bob," "Mom," and "pop" were accepted as correct when participants viewed the articulation of "mop."

Table 3: Words presented in the visual only condition

Bilabial/labiodental onset		Alveolar onset	
one-syllable words	two-syllable words	one-syllable words	two-syllable words
boy	pumpkin	dog	necklace
mop	mailman	tree	donkey
farm	window	lamb	sweater
beach	balloon	knife	zipper
woods	flower	scarf	teacher

2.4 Sequence of Testing Sessions

All testing occurred over 3 sessions administered on 3 different days. One of the SIN task conditions (either A or AV) was administered during the first session, the audiovisual matching task and the lip-reading task – during the second session (with the lip-reading task always preceding the audiovisual matching task), and the second SIN task condition – during the third session. Because the same words were used in the audiovisual matching task and in the SIN task, most participants' sessions were separated by at least 7 days to minimize the possible effect of stimulus repetition.

2.5 Statistical Analyses

2.5.1 Behavioral and ERP measures

One-way ANOVA tests were used to compare group means on all screening tests. The homogeneity of variances across groups was evaluated with the Levene statistic. When variances differed, the Brown-Forsythe correction was applied. In all such cases, the corrected degrees of freedom and *p*-value are reported.

Repeated-measures ANOVAs were used to determine whether groups differed in the number of correct responses, incorrect responses, misses, and in reaction time during the audiovisual matching task and to evaluate whether the SIN accuracy was higher in the AV compared to the A condition. Repeated-measures ANOVAs were also used to evaluate ERP components. When omnibus ANOVA analysis produced a significant interaction, it was further analyzed with step-down ANOVAs, with factors specific to any given interaction. When the assumption of sphericity was violated, we used the Greenhouse-Geisser adjusted *p*-values to determine significance. Effect sizes, indexed by the partial eta squared statistic (η_p^2), are reported for all significant repeated-measures ANOVA results.

2.5.2 Regressions

Multiple regression analyses aimed to determine how the relationship between ERP measures of visual articulatory processing (i.e., N400 and LPC) and gains during audiovisual as compared to auditory SIN perception changes with age. A hierarchical forced-entry regression model was constructed for each group of participants, in which ERP measures were entered as predictors and the SIN gain in the audiovisual SIN task as an outcome. Based on our earlier studies, we knew that at least in adults, N400 was related to the SIN improvement (Kaganovich et al., 2016b). Therefore, the N400 mean amplitude was entered in the first step of the model

construction, with the LPC entered in the second step. The N400 and LPC measures entered into

regressions were their mean differences between congruent and incongruent trials averaged

across all sites used for statistical analyses. To screen for outliers, we used the standardized

DFBeta function in the SPSS Statistics program. Cases with the standardized DFBeta values over

1 have a significant influence over the regression model and are considered outliers (Field,

2013). No such cases were detected. Variance inflation factor (VIF) screened for

multicollinearity. None of the VIF numbers exceeded 10, with the average VIF being 1.6,

suggesting that multicollinearity was not a significant factor (Field, 2013).

CHAPTER 3. RESULTS

3.1 Experiment 1 – Audiovisual Matching Task

3.1.1 Behavioral Results

Behavioral performance on the audiovisual matching task is summarized in Table 4. Across all groups, *percentage correct* (percentage of trials in which the participant selected the right response) did not differ between congruent (mean=94.50%) and incongruent (mean=93.80%) articulations, $F(1, 39)=0.45$, $p=.50$, $\eta_p^2=.01$. While all groups obtained a high percentage correct for this task, the groups still differed $F(2, 39)=9.34$, $p<.01$, $\eta_p^2=.32$. Importantly, there was no group by congruency interaction, $F(2, 39)=0.58$, $p=.56$, $\eta_p^2=.03$. Across both conditions, pair-wise comparisons showed that the 8-9-year old group performed significantly worse (mean=91.29%) than adults (mean=96.96%), $p<.01$, and marginally worse than the 11-12-year old group (mean=94.20%), $p=.10$. Older children did not differ from adults, $p=.13$.

Across all groups, the *percentage missed* (percentage of trials on which the participant did not respond within the expected time frame) did not differ between congruent articulations (mean=1.63%) and incongruent articulations (mean=1.98%), $F(1,39)=1.25$, $p=.27$, $\eta_p^2=.03$. The main effect of group was not significant, $F(2,39)=2.27$, $p=.12$, $\eta_p^2=.1$, nor was there a group by congruency interaction, $F(2,39)=0.71$, $p=.50$, $\eta_p^2=.04$.

Across all groups, comparison of *reaction times* showed that participants were significantly faster at responding to congruent (mean=711.87 ms) than to incongruent articulations (mean=737.01 ms), $F(1,39)=4.83$, $p=.03$, $\eta_p^2=.11$. The main effect of group was significant, $F(2,39)=3.55$, $p=.04$, $\eta_p^2=.15$. Across both conditions, the adult group (mean=629.25

ms) was marginally faster than the 8-9-year old group (mean=769.61 ms), p=.09, and the 11-12-year old group (mean=774.31 ms), p=.07. Older children did not differ from younger children in their reaction time, p=1.00.

Table 4: Performance on the audiovisual matching task.

	8-9 year old	11-12 year olds	Adults
Percentage Correct			
Congruent	92.41 (5.01)	94.05 (3.36)	97.03 (3.02)
Incongruent	90.17 (7.04)	94.35 (5.69)	96.89 (2.56)
Percentage Missed			
Congruent	2.98 (2.80)	1.04 (1.58)	0.88 (1.54)
Incongruent	2.83 (3.42)	1.79 (2.56)	1.34 (3.02)
Reaction Time (ms)			
Congruent	766.11 (142.88)	762.85 (97.88)	606.64 (216.97)
Incongruent	773.41 (147.48)	785.77 (78.53)	651.86 (253.00)

Note. Numbers in parenthesis are standard deviation of the mean.

3.1.2 ERP Results

Figure 3 shows overlays of the ERP waveforms elicited in the congruent and incongruent conditions for each group. Figure 4 shows a difference waveform between incongruent and congruent conditions for all groups. Additionally, mean values and standard errors for the N400 and LPC mean amplitudes for each group are detailed in Figure 5.

All groups displayed the *N400 component* following incongruent articulations. Mean amplitude and latency were compared based on a subtraction of the congruent condition from the incongruent condition. In initial analyses, the effect of hemisphere was not significant, $F(1,39)$=0.04, p=.84, η_{p}^{2}<.01, nor was there a group by hemisphere interaction, $F(2,39)$=0.49, p=.62, η_{p}^{2}=.03. Therefore, data from electrodes over both hemispheres as well as from the midline Pz site were combined for further analyses. In absolute terms, the 8-9-year old group had the greatest N400 mean amplitude, followed by the 11-12-year old group. The adults had the smallest N400. The main effect of group was significant $F(2,39)$=3.22, p=.05, η_{p}^{2}=.14. Bonferroni-corrected pairwise comparisons revealed that the 8-9-year old group had a

significantly larger mean amplitude than adults, $p=.06$. The 11-12 group did not differ

significantly from either the 8-9 group, $p=1.00$, or the adult group, $p=.25$.

Analysis of the N400 mean latency showed that the effect of hemisphere was not

significant, $F(1,39)=3.66$, $p=.06$, $\eta_p^2=.09$, nor was there a group by hemisphere interaction,

$F(2,39)=2.58$, $p=.09$, $\eta_p^2=.12$ Therefore, data from electrodes over both hemispheres as well as

from the midline Pz site were combined for further analyses. For N400 mean latency, the main

effect of group was not significant $F(2,39)=1.67$, $p=.20$, $\eta_p^2=.08$.

All groups displayed the *LPC component* following congruent articulations. Mean

amplitude and latency were compared based on a subtraction of the incongruent condition from

the congruent condition. In initial analyses of the LPC amplitude, there was no main effect of

hemisphere, $F(1, 39)= 1.41$, $p=.24$, $\eta_p^2=.04$, and no hemisphere by group interaction,

$F(2,39)=0.73$, $p=.49$, $\eta_p^2=.04$; therefore, data from electrodes over both hemispheres as well as

from the midline Pz and Oz sites were combined for further analyses. The main effect of group

was significant, $F(2,39)=6.60$, $p<.01$, $\eta_p^2=.25$. Bonferroni-corrected pairwise comparisons

revealed that the 8-9-year old group had a significantly larger mean amplitude than both the 11-

12-year old group, $p=.03$, and the adult group, $p<.01$. The 11-12-year old group and the adult

group did not differ, $p=1$. In initial analyses of the LPC mean latency, there was a main effect of

hemisphere, $F(1, 39)= 4.23$, $p=.05$, $\eta_p^2 =.10$, with the right hemisphere (mean=1.21s) having a

longer latency than the left hemisphere (mean=1.19s). There was no hemisphere by group

interaction, $F(2,39)=1.71$, $p=.20$, $\eta_p^2=.08$. Additionally, the main effect of group was not

significant $F(2,39)=0.22$, $p=.81$, $\eta_p^2=.01$.
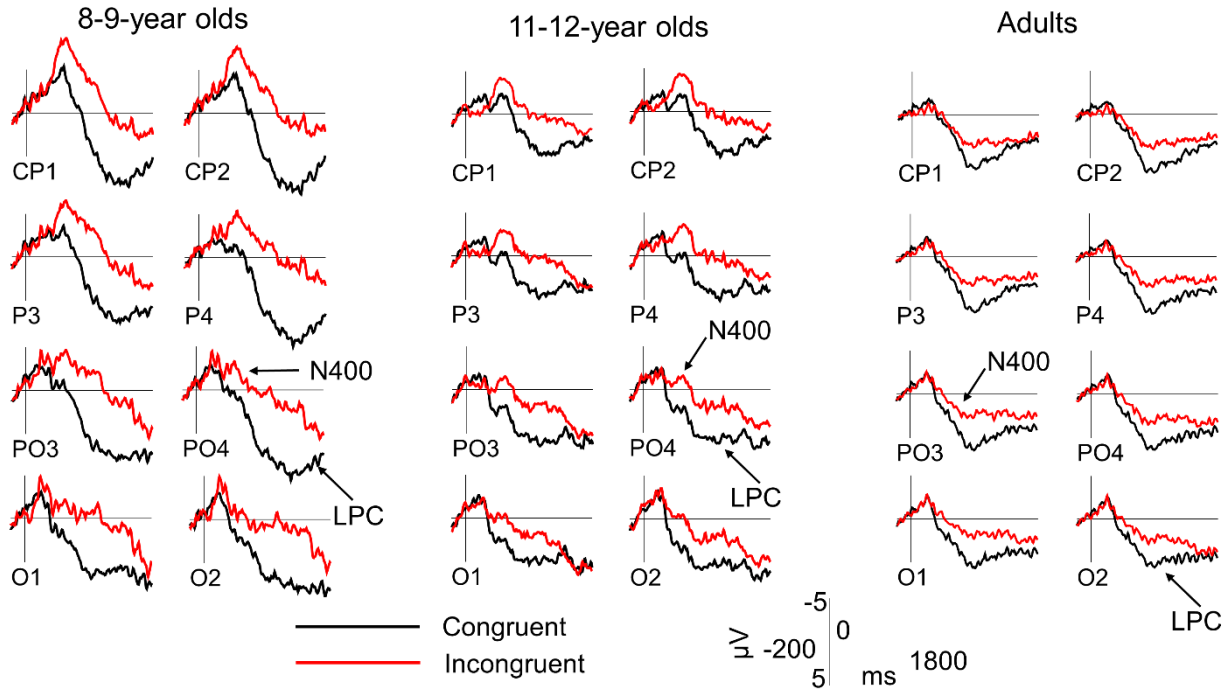
Figure 3: ERPs Elicited from Conguent and Incongruent Articulations. Each section (8-9-year olds, 11-12-year olds, and Adults) shows an overlap between grand average waveforms elicited by congruent and incongruent articulations by group. In the 8-9-year old group and the 11-12-year old group, the N400 and LPC components are marked on the PO4 site. For the adults, the N400 and LPC components are marked on the PO3 and O2 sites, respectively. Negative is plotted up. Time 0 indexes the onset of articulatory movements.
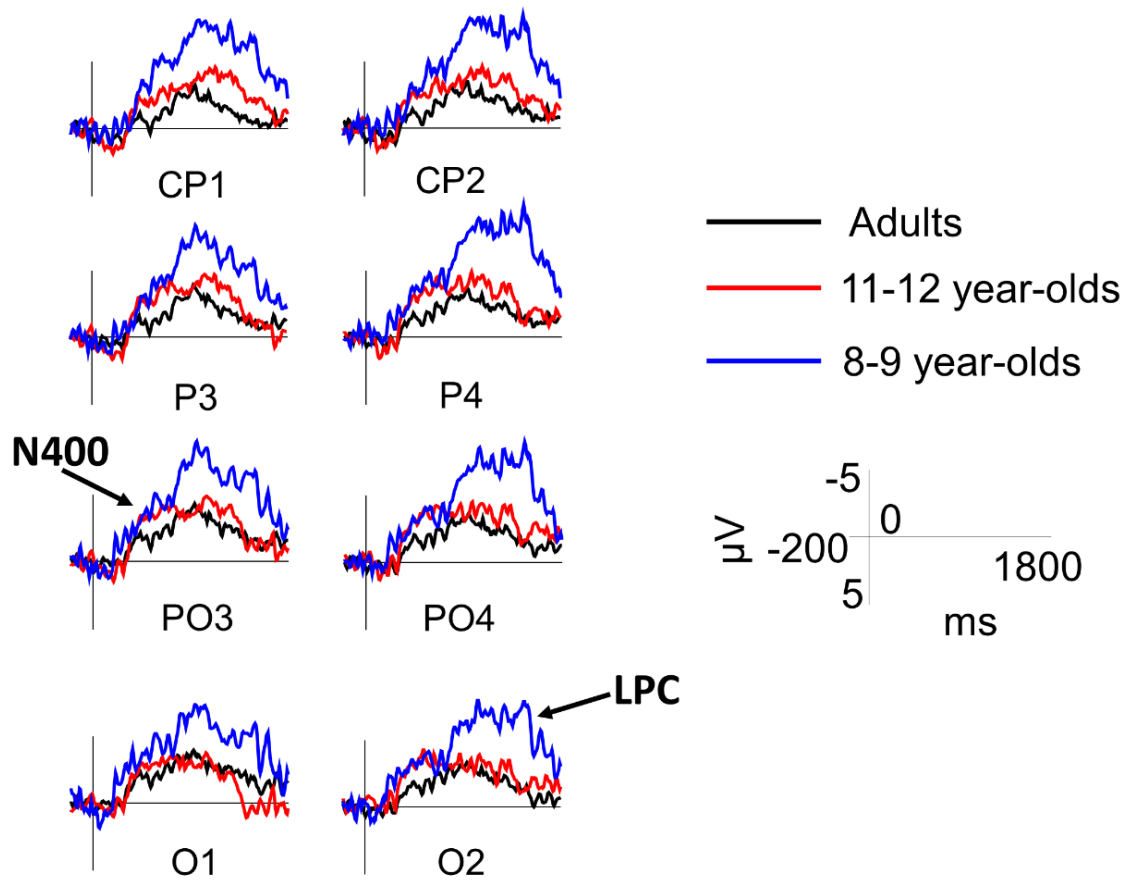
Figure 4: ERP Difference Waves: Incongruent Minus Congruent. Figure shows a difference waveform produced by subtracting the ERPs elicited by incongruent articulations from the ERPs elicited by congruent articulations for each group. The N400 and LPC components are marked on the PO3 and O2 sites, respectively. Negative is plotted up. Time 0 indexes the onset of articulatory movements.
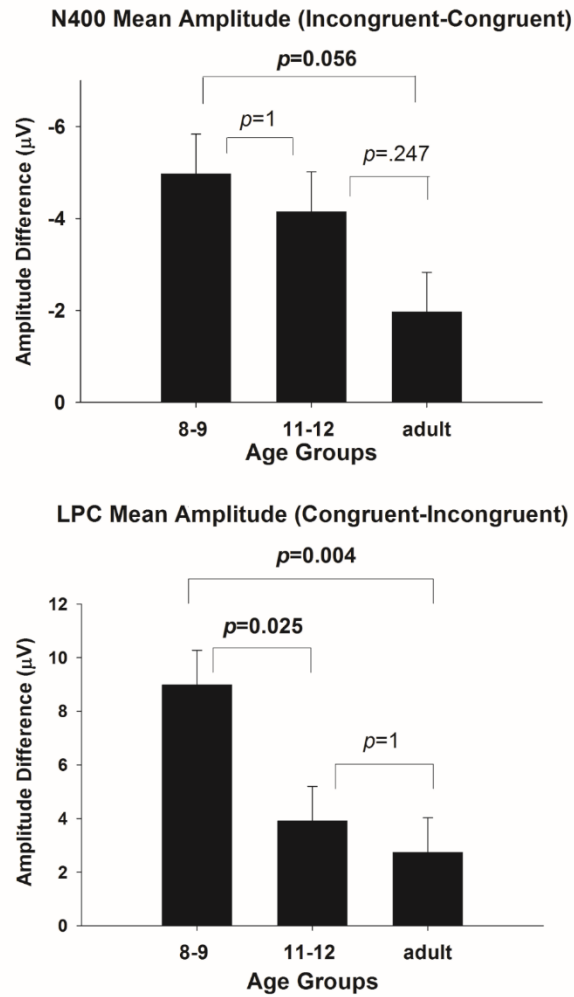
Figure 5: N400 and LPC Mean Amplitudes across Groups. Error bars show standard errors of the mean. Both N400 and LPC amplitude differences measures are based on a subtraction between waveforms from the two conditions: incongruent-congruent for N400 and congruent-incongruent for LPC. P-values show results of Bonferroni-corrected pairwise comparisons between groups.

3.2 Experiment 2 – Word Identification Tasks

Table 5 shows means and standard deviations for the auditory only (A) and audiovisual (AV) conditions in SIN task, the gain from seeing the talker's face (AV-A), and the lip-reading task (V). For the lip-reading (V) task, the groups did not differ in the percentage of words correctly identified, $F(2,14)= 1.71$, $p=.19$.

For the SIN tasks (A and AV), the groups did significantly better in the AV condition (mean=80.66%) than in the A condition (mean=50.23%), $F(1,39)=180.19$, $p<.01$, $\eta_p^2=.82$. Across

both conditions, the groups differed in the percentage of words they were able to correctly identify, $F(2,39)=64.64$, $p<.01$, $\eta_{p^2}=.77$. Bonferroni-corrected pairwise comparisons revealed that the 8-9-year olds were outperformed by both the 11-12-year old group, $p<.01$, and the adult group, $p<.01$. The adults also performed better than the 11-12-year old group, $p<.01$. The groups all gained from seeing the talker's face during speech-in-noise tasks, but did not differ in the amount of this gain, $F(2, 41)= 1.46$, $p= .25$.

Table 5: Group performance on word identification tasks

	8-9 year olds	11-12 year olds	Adults
SIN Task			
A	38.38 (9.48)	47.72 (9.90)	61.78 (5.80)
AV	72.01 (3.68)	82.16 (5.05)	90.61 (3.96)
AV-A	33.64 (10.94)	34.44 (9.25)	28.83 (7.74)
Lip-reading Task			
V	32.14 (18.99)	37.50 (18.89)	44.29 (13.85)

Note. AV-A indexes the amount of gain from seeing the talker's face. Numbers for A, AV, and V are the percentage of words that were correctly identified. A auditory only, AV audiovisual, V visual only (lipreading). Standard deviations are shown next to each mean.

3.3 Regressions

Table 6 displays regression results. Figure 6 shows relationships between the N400 and LPC ERP components and SIN improvement in the presence of the talker's face.

In the younger group of children, neither ERP component was a strong predictor of audiovisual SIN improvement. Although the addition of LPC in the second step had a marginally significant influence on the model ($p=.09$), the correlation coefficient's 95% confidence interval spanned the 0 point, suggesting that LPC was a poor predictor of SIN performance. In the older group of children, the N400 component entered in step 1 had a negligible influence on the model. However, the addition of the LPC component in step 2 had a significant influence ($p=.03$), accounting for 37% of variance in the SIN performance. Importantly, the LPC's correlation coefficient was negative, suggesting that larger LPC was associated with smaller improvement

on the audiovisual version of SIN. Finally, in adults, the amplitude of N400 was a strong

predictor of audiovisual SIN gain (p=.02). It accounted for 40% of variance in the SIN

improvement. Although the addition of LPC in step 2 increased the R^2 value from .4 to .45, the

LPC's influence on the model was not significant (p=.33).

Table 6: Regression results.

	B (95% CI)	*SE of B*	*Std Beta*	*p*	*R^2*
8-9-year olds					
Step 1					
Constant	36.5 (25.9, 47.2)	4.89		.00	.04
N400	0.58 (-1.12, 2.28)	0.78	.21	.47	
Step 2					
Constant	41.4 (30, 52.8)	5.2		.00	
N400	-0.28 (-2.15, 1.6)	0.85	-.1	.75	.27
LPC	-1.02 (-2.23, 0.19)	0.55	-.57	**.09**	
11-12-year olds					
Step 1					
Constant	32.9 (23.1, 42.7)	4.5		.00	
N400	-0.37 (-2.32, 1.58)	0.9	-.12	.68	.01
Step 2					
Constant	3.38 (25.1, 41.67)	3.76		.00	
N400	-1.98(-4.14, 0.19)	0.98	-.63	**.07**	.37
LPC	-1.83 (-3.44, -0.22)	0.73	-.79	**.03**	
Adults					
Step 1					
Constant	25.2 (20.6, 29.8)	2.1		.00	
N400	-1.84 (-3.3, -0.4)	0.65	-.63	**.02**	.4
Step 2					
Constant	25.64 (20.92, 30.37)	2.14		.00	
N400	-2.43 (-4.34, -0.51)	0.87	-.83	**.02**	.45
LPC	-0.58 (-1.82, 0.67)	0.56	-.3	.33	

Note. For all analyses, ERP measures were entered as predictor variables while the SIN improvement in the presence of the talker's face as an outcome variable. Both N400 and LPC measures are based on the difference waveforms: incongruent-congruent for N400 and congruent-incongruent for LPC. All *p*-values below .1 are bolded for the ease of reference.
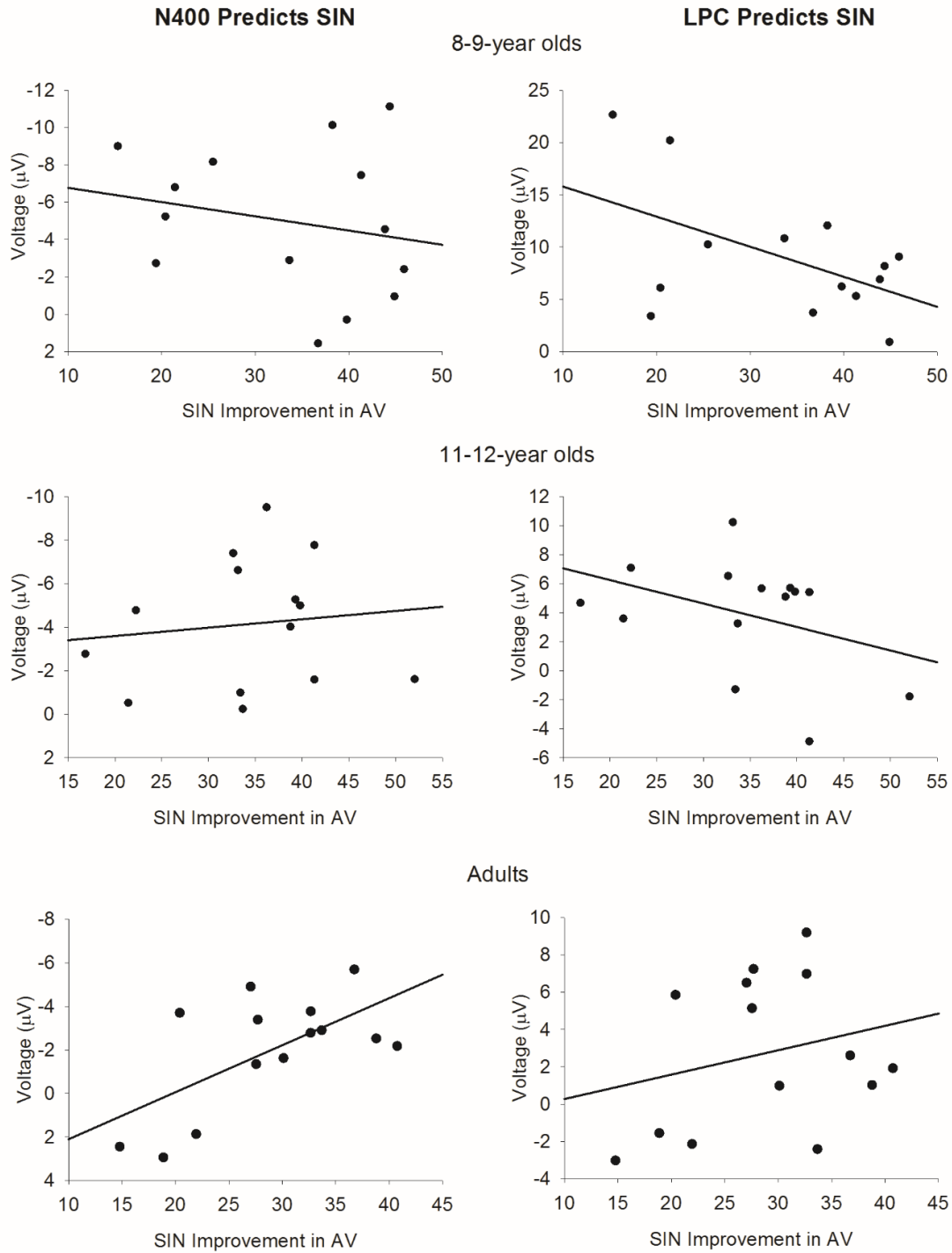
Figure 6: Relationships between the N400 and LPC ERP Components and SIN Improvement in the Presence of the Talker's Face. For each group, the relationship between the N400 and SIN improvement is shown on the left, while the relationship between the LPC and SIN improvement is shown on the right. Both N400 and LPC resuls are based on the difference waveforms: incongruent-congruent for N400 and congruent-incongruent for LPC. The SIN improvement measure is calculated by subtracting the A condition from the AV condition in the SIN task.

CHAPTER 4. DISCUSSION

In this study, we examined children and adults' responses to visually-perceived articulatory movements that were either congruent or incongruent with an auditory word presented beforehand. Overall, our results suggest that while audiovisual word matching as indexed by the LPC becomes adultlike by the age of 11 or 12, phoneme-to-articulation mapping as reflected by the N400 is not fully matured until later in development. In addition, we have evidence that the relationship of the N400 and the LPC to a real-life skill, such as speech-in-noise perception, changes throughout the lifespan. Specifically, the relation of the LPC to the amount of SIN improvement in the presence of the talker's face is stronger earlier in childhood while the relation of the N400 is stronger during later school years and adulthood.

To understand the significance of these results it is first important to understand the development of lexical representations throughout childhood. To complete the Audiovisual Matching Task successfully, participants must first produce visual mental representations of the auditory word to then compare it with the perceived visual articulations. Once these are compared, the participants decide whether the expected and the observed articulation depicted the same lexical item. In children, early representations of the lexical system are holistic in nature and focus on general structure. Later on, as a result of experience with speech perception and other factors, lexical representations become more detailed (for a summary, see Metsala & Walley, 1998). These representations contain information about smaller units of speech, such as syllables and phonemes. The development of these representations is thought to depend partially on a child's vocabulary growth, and children are estimated to gain understanding of about 25,000 words from grade 3 to grade 8 (Segbers & Schroeder, 2017). In other words, the ages of our

groups encompass a period where the lexical system is still undergoing changes. Our results are indicative of these developmental changes between groups. As stated previously, the N400 is sensitive to information on a phonological level, meaning that the N400 response is contingent on the participant having a more detailed mental representation of a visually perceived word. In contrast, the LPC is associated with the recognition of a repeated word on a whole-word level during congruent trials, which does not necessarily require a detailed mental representation for comparison. Our results show that the development of the N400, which relies on more fine-grained detail, occurs later than the LPC, which relies on the whole-word level.

Not only are the developmental trajectories of these components distinct, we have also shown that younger children's LPC amplitude is more strongly associated with SIN improvement while older participant's N400 amplitude is more strongly associated with SIN improvement. This result is in accordance with what we know from the development of lexical representations in children. Since children have more holistic representations of words, instead of more detailed, phonological representations, it follows that a component related to the recognition of a phonological mismatch, the N400, would not show sufficient maturity to relate to a real-life skill, in this case SIN improvement.

Another important finding of our study is that cognitive ERP components, such as the N400 and the LPC, have a surprisingly extended developmental period. In our study, the child groups had, on average, high accuracy during the Audiovisual Matching Task when asked to identify mismatches, and they had roughly equivalent gains in audiovisual processing during the SIN tasks. However, their ERP responses show that they are remarkably immature compared to their level of performance. In general, sufficiently accurate behavioral performance should not be

equated to having "adult-like" brain responses, since a wide range of maturity can lead to similar behavioral results in some tasks.

Like any study, ours has its limitations that may be addressed by future research. To explore the development of the N400 and LPC throughout childhood in greater depth, it would be beneficial to widen the scope of this research. This study was relatively small, with 14 participants in each group. Adding more participants in all groups could yield stronger results and more reliable patterns of development. Also, it would be beneficial to encompass a wider age range within the course of development. With the 8-9 year old and the 11-12 year olds, we were able to capture important evidence of maturation during pre-adolescence. However, given the amount of behavioral and neural maturation related to audiovisual speech perception that takes place outside of this age range, we have presented only a snapshot of the developmental progression.

Our results represent an important step in tracking the development of task-specific neural responses related to audiovisual integration for speech. Audiovisual speech perception is an important skill for children to develop, and, therefore, it is essential to explore how the relationship between behavioral and neural responses changes with age. Although our study has explored only a small window of that development, it can be used as a stepping stone for further research. In addition to exploring a wider age range, further manipulation of our paradigm could also advance our knowledge of how the position of audiovisual information within a word may affect its processing. Including more subtle incongruencies at varying locations within a word, apart from the onset, would be able to further tease apart individual differences in responses to audiovisual speech.

Not only is this exploration important for typical development, it also has specific clinical implications. There is evidence supporting the idea that audiovisual integration may be impaired in a number of specific populations. Children with Attention-Deficit Hyperactivity Disorder (Panagiotidi, Overton, & Stafford, 2017), Autism Spectrum Disorder (Foxe et al., 2013), Schizophrenia (Pearl et al., 2009), and Dyslexia (Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009) have all shown deficits in audiovisual processing. In order to establish a more nuanced perspective on differences in those populations, it is essential to better understand the development of audiovisual integration-related processes in typical children.

REFERENCES

American Electroencephalographic Society. (1994). Guideline thirteen: Guidelines for standard electrode placement nomenclature. *Journal of Clinical Neurophysiology, 11*, 111-113.

Annaz, D., Karmiloff-Smith, A., & Thomas, M. S. C. (2007). The importance of tracing developmental trajectories for clinical child neuropsychology.

BioSemi. (2013). Active Electrodes. Retrieved from http://www.biosemi.com/active_electrode.htm

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology, 19*(6), 503-508.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (version 5.3) [Computer program]. Retrieved from http://www.praat.org (Version 5.1).

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence* (4th ed.). Austin, Texas: Pro-Ed: An International Pubilsher.

Cohen, M. S. (2008). Handedness Questionnaire.Retrieved from http://www.brainmapping.org/shared/Edinburgh.php#

Conners, K. C. (1997). *Conners' Rating Scales - Revised*. North Tonawanda, NY: MHS.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1136-1146.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th Ed.): Pearson.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America, 66*(3). doi:10.1121/1.383691

Fenson, L., Marchman, V., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories (CDI) Words and Sentences*: Brookes Publishing Co.

Field, A. (2013). *Discovering statistics using SPSS* (4 ed.). Washington, DC: Sage.

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H.-P., Russo, N. N., Blanco, D., . . . Ross, L. A. (2013). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during adolescence. *Cerebral Cortex*. doi:10.1093/cercor/bht213

Jetha, M. K., & Segalowitz, S. (2012). Structural Brain Development in Late Childhood, Adolescence, and Early Adulthood. *Adolescent Brain Development: Implications for Behavior*, 1-18. doi:10.1016/B978-0-12-397916-2.00001-3

Kaganovich, N., Schumaker, J., & Rowland, C. (2016a). Atypical audiovisual word processing in school-age children with a history of specific language impairment: an event-related potential study. *Journal of Neurodevelopmental Disorders, 8*(33). doi:10.1186/s11689-016-9168-3

Kaganovich, N., Schumaker, J., & Rowland, C. (2016b). Matching heard and seen speech: An ERP study of audiovisual word recognition. *Brain and Language, 157-158*, 14-24.

Kirk, K. I., Pisoni, D. B., & Lachs, L. (2002). Audiovisual integration of speech by children and adults with cochlear implants. *Proceedings: ICSLP International Conference on Spoken Language Processing, 2002*, 1689-1692.

Martin, N., & Brownel, R. (2005). *Test of Auditory Processing Skills* (3rd ed.). Novato, California: Academic Therapy Publications.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. *Word recognition in beginning literacy*, 89-120.

Metting van Rijn, A. C., Kuiper, A. P., Dankers, T. E., & Grimbergen, C., A. (1996). *Low-cost active electrode improves the resolution in biopotential recordings.* Paper presented at the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam, The Netherlands.

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences, 39*(8), 567-579. doi:10.1016/j.tins.2016.05.003

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*, 4-12.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Panagiotidi, M., Overton, P. G., & Stafford, T. (2017). Multisensory integration and ADHD-like traits: Evidence for an abnormal temporal integration window in ADHD. *Acta psychologica, 181*, 10-17. doi:10.1016/j.actpsy.2017.10.001

Papso, C. F., & Blood, I. M. (1989). Word recognition skills of children and adults in background noise. *Ear and Hearing, 10*(4), 235-236.

Pearl, D., Yodashkin-Porat, D., Katz, N., Valevski, A., Aizenberg, D., Sigler, M., . . . Kikinzon, L. (2009). Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. *Comprehensive Psychiatry, 50*(2), 186-192.

Pflieger, M. E. (2001). *Theory of a spatial filter for removing ocular artifacts with preservation of EEG.* Paper presented at the EMSE Workshop, Princeton University. http://www.sourcesignal.com/SpFilt_Ocular_Artifact.pdf

Pimperton, H., Ralph-Lewis, A., & MacSweeney, M. (2017). Speechreading in Deaf Adults with Cochlear Implants: Evidence for Perceptual Compensation. *Frontiers in Psychology, 8*(106). doi:10.3389/fpsyg.2017.00106

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research, 1*, 73-86.

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *The European Journal of Neuroscience, 33*(12), 2329-2337. doi:10.1111/j.1460-9568.2011.07685.x

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research, 1188*, 87-89. doi:10.1016/j.brainres.2007.10.049

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

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*(6), 251-257.

Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory & Cognition, 15*(6), 473-481.

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood Autism Rating Scale* (2nd ed.): Western Psychological Services.

Schorr, E. A., Fox, N. A., Van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America, 102*(51), 18748-18750. doi:10.1073/pnas.0508862102

Segbers, J., & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing, 34*(3), 297-320.

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

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *CELF4: Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: Pearson Clinical Assessment.

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging, 36*, 283-291.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temproal cues facilitate detection. *Neuropsychologia, 63*, 43-50.

Thompson, L. A., & Malloy, D. (2004). Attention resources and visible speech encoding in older and younger adults. *Experimental Aging Research, 30*(3), 241-252.

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception and Psychophysics, 60*(6), 926-940.

WMA Declaration of Hesinki: ethical principles for medical research involving human subjects. (1964).

APPENDIX

Appendix. The pairing of auditory words and silent visual articulations.
Note that articulations that are congruent (i.e., match the preceding auditory word) in List A are
incongruent (i.e., do not match the preceding auditory word) in List B.

	List A		List B	
	Auditory Word	**Silent Visual Articulation**	**Auditory Word**	**Silent Visual Articulation**
1	shower	shower	candy	donkey
2	tree	lamb	cat	cat
3	jello	jello	jello	monkey
4	cat	girl	egg	egg
5	egg	pool	donut	bottle
6	donut	donut	zipper	present
7	zipper	zipper	donkey	candy
8	donkey	donkey	shirt	shirt
9	grapes	grapes	grapes	farm
10	police	apple	police	police
11	truck	belt	apple	apple
12	apple	police	truck	truck
13	monkey	monkey	monkey	jello
14	sandwich	mailman	sandwich	sandwich
15	car	car	car	fish
16	turtle	turtle	turtle	popcorn
17	squirrel	squirrel	squirrel	pretzel
18	window	window	window	sandbox
19	sled	bird	sled	sled

	List A		List B	
	Auditory Word	**Silent Visual Articulation**	**Auditory Word**	**Silent Visual Articulation**
20	necklace	necklace	bread	duck
21	water	water	water	carrot
22	sink	sink	sink	mop
23	paint	paint	paint	woods
24	pretzel	pretzel	pretzel	squirrel
25	nail	peas	nail	nail
26	bird	sled	bird	bird
27	corn	corn	corn	frog
28	couch	couch	couch	moose
29	farm	farm	farm	grapes
30	airplane	pumpkin	airplane	airplane
31	popcorn	popcorn	popcorn	turtle
32	penguin	doctor	penguin	penguin
33	knife	mouth	mouth	mouth
34	arm	horse	arm	arm
35	bed	ear	bed	bed
36	present	present	present	zipper
37	sandbox	sandbox	sandbox	window
38	mop	mop	mop	sink
39	mailman	sandwich	mailman	mailman
40	lamb	tree	shower	necklace
41	candy	candy	duck	bread
42	scissors	balloon	scissors	scissors
43	pool	egg	pool	pool
44	bee	bee	bee	eye
45	chair	boat	chair	chair

	List A		List B	
	Auditory Word	**Silent Visual Articulation**	**Auditory Word**	**Silent Visual Articulation**
46	cake	ball	cake	cake
47	boy	boy	boy	dog
48	sprinkler	muffin	sprinkler	sprinkler
49	elephant	elephant	elephant	teddy bear
50	comb	beach	comb	comb
51	jar	purse	jar	jar
52	horse	arm	horse	horse
53	sweater	sweater	sweater	picture
54	moose	moose	moose	couch
55	muffin	sprinkler	muffin	muffin
56	ear	bed	ear	ear
57	toys	toys	toys	bus
58	bus	bus	carrot	water
59	carrot	carrot	teacher	buttons
60	teacher	teacher	hammer	hammer
61	hammer	pizza	bus	toys
62	frog	frog	frog	corn
63	shirt	foot	necklace	shower
64	buttons	buttons	buttons	teacher
65	ball	cake	ball	ball
66	beach	comb	beach	beach
67	girl	cat	girl	girl
68	mouth	knife	knife	knife
69	peas	nail	peas	peas
70	woods	woods	woods	paint
71	picture	picture	picture	sweater

	List A		List B	
	Auditory Word	**Silent Visual Articulation**	**Auditory Word**	**Silent Visual Articulation**
72	purse	jar	purse	purse
73	belt	truck	belt	belt
74	wolf	wolf	wolf	house
75	scarf	scarf	scarf	broom
76	teddy bear	teddy bear	teddy bear	elephant
77	house	house	house	wolf
78	eye	eye	eye	bee
79	dog	dog	dog	boy
80	flower	orange	flower	flower
81	doctor	penguin	doctor	doctor
82	foot	shirt	foot	foot
83	broom	broom	broom	scarf
84	tractor	pencil	tractor	tractor
85	circus	money	circus	circus
86	balloon	scissors	balloon	balloon
87	orange	flower	orange	orange
88	pencil	tractor	pencil	pencil
89	pumpkin	airplane	pumpkin	pumpkin
90	bread	bread	lamb	lamb
91	money	circus	money	money
92	bottle	bottle	bottle	donut
93	boat	chair	boat	boat
94	pizza	hammer	pizza	pizza
95	fish	fish	fish	car
96	duck	duck	tree	tree