The Thesis Committee for Rachel Denise Reetzke
Certifies that this is the approved version of the following thesis:

Developmental and Cultural Factors of Audiovisual Speech Perception in Noise

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Li Sheng

Co-Supervisor: ______________________________
Bharath Chandrasekaran

Developmental and Cultural Factors of Audiovisual Speech Perception in Noise

by

Rachel Denise Reetzke, B.S.

Thesis

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts

The University of Texas at Austin
May 2014

Dedication

To my family, friends, and colleagues for their advice, encouragement, love and support throughout this project.

"When the eye is unobstructed, the result is sight. When the ear is unobstructed, the result is hearing. When the mind is unobstructed, the result is truth. When the heart is unobstructed, the result is joy and love." –Anthony DeMello

Acknowledgements

I am grateful to the children, parents, university students, and professors who made this study possible. I would like to thank Dr. Li Sheng and Dr. Bharath Chandrasekaran, my supervisors, for sharing their expertise and the use of the SoundBrain Laboratory and Language Learning and Bilingualism Laboratory equipment, space, and resources. Thank you to Zilong Xie and Boji Lam for assistance with the statistical analysis and graphs. I would additionally like to thank my primary research assistants Rachel Tessmer and Nicole Tsao for their assistance throughout the entire project. Finally, I am grateful to Kathryn Gay, Katie Keith, Robyn Ward and Hannah Humphrey for assistance with accuracy scoring, calculation, and inter-rater reliability.

Abstract

Developmental and Cultural Factors of Audiovisual Speech Perception in Noise

Rachel Denise Reetzke, M.A.
The University of Texas at Austin, 2014

Supervisors: Li Sheng and Bharath Chandrasekaran

The aim of this project is two-fold: 1) to investigate developmental differences in intelligibility gains from visual cues in speech perception-in-noise, and 2) to examine how different types of maskers modulate visual enhancement across age groups. A secondary aim of this project is to investigate whether or not bilingualism differentially modulates audiovisual integration during speech in noise tasks. To that end, both child and adult, monolingual and bilingual participants completed speech perception in noise tasks through three within-subject variables: (1) masker type: pink noise or two-talker babble, (2) modality: audio-only (AO) and audiovisual (AV), and (3) Signal-to-noise ratio (SNR): 0 dB, -4 dB, -8 dB, -12 dB, and -16 dB. The findings revealed that, although both children and adults benefited from visual cues in speech-in-noise tasks, adults showed greater benefit at lower SNRs. Moreover, although child monolingual and bilingual participants performed comparably across all conditions, monolingual adults outperformed simultaneous bilingual adult participants. These results may indicate that the divergent use of visual cues in speech perception between bilingual and monolingual speakers occurs later in development.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

According to a 2013 US Census report, there are approximately 83 million students attending elementary school through university in the United States (Davis & Bauman, 2013). With the advancement of a global society, this significant portion of our population is far from homogenous, containing an amalgam of ages, cultures, and abilities. Research over the past several years has indicated that classroom acoustics significantly impact a student's academic achievement (e.g. Hetu, Truchon-Gagnon, & Bilodeau, 1990; Crandell & Smaldino, 1996; Picard & Bradley, 2001; Crandell & Smaldino, 1996; Picard & Bradley, 2001). For example, Hetu et al. (1990) found that younger children are more distracted by noise when compared to older children in the classroom environment, and more recently, Riley & McGregor (2012) found that classroom noise limits expressive vocabulary growth in school age children. The detrimental impact of classroom acoustics is found throughout a student's academic career, as studies reveal that adverse listening conditions negatively impact university-age students as well (Hodgson, 2002; for a review, see Picard & Bradley, 2001).

Before understanding how adverse listening conditions modulate learning in the classroom, the modalities which students utilize to perceive speech in the environment must first be understood. In the past, speech perception was largely studied as an auditory unimodal phenomenon. However, a plethora of evidence over the past few decades has demonstrated that speech perception is substantially influenced by visual input (e.g. Sekiyama & Burnham, 2008; for a review, see Woodhouse, Hickson, & Dodd, 2009). Unfortunately, evidence thus far does not converge on a conclusion regarding how and when audiovisual integration processes develop across the lifespan (Navarra, Yeung, Werker, & Soto-Faraco, 2012). Therefore, in order to better understand the developmental trajectory of the auditory and visual integration, the utilization of modalities should be observed in both child and adult participants' speech perception performance in adverse listening conditions.

How do we test speech perception-in-noise? Unfortunately, the majority of routine clinical practice does not assess an individual's ability to understand speech in adverse listening conditions (Picard & Bradley, 2001). In turn, the evidence that we have regarding speech in noise tasks is mainly auditory-only speech perception, rather than multisensory audiovisual speech perception (Picard & Bradley, 2001; Riley & McGregor, 2012). Therefore, current investigations and available findings of speech perception-in-noise have mostly focused on the listener's speech perception in a restricted range of conditions, dissimilar to the everyday listening environment.

The difficulty associated with understanding speech in suboptimal environments is typically categorized into one of two types of adverse listening condition categories: energetic masking and informational masking (Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001). Energetic masking occurs when competing signals overlap in time and frequency, which in turn causes one or more of the signals to be perceived as less audible. In contrast, informational masking categorizes adverse listening conditions where the target and masker signals are clearly audible but the listener is unable to segregate the elements of the target signal from the features of the similar-sounding distracters.

Few studies to date (e.g. Ross et al., 2011) have focused on the developmental aspects of audiovisual speech perception-in-noise, leaving a gap in knowledge regarding the specific developmental trajectory of these salient modalities. Sumby & Pollack (1954) pioneered one of the first studies to investigate an individual's utilization of visual cues during speech perception-in-noise tasks. This study indicated that when an individual is able to see a speaker's face along with the auditory signal, speech intelligibility increases in comparison to auditory signal only performance. However, Sumby & Pollack used a restricted set of word stimuli that were presented to subjects before and during the experiments. Moreover, they designed their experiments to simulate only one type of adverse listening condition in the form of energetic masking.

While Sumby & Pollack provided novel insight into the visual modality and its benefit to speech intelligibility in adverse listening conditions, this study also prompted a protocol for restricted speech in noise experiments. Studies to date typically present limited speech stimuli, such as a single sound (e.g. Schwartz, Berthommier, & Savariaux, 2004) or a single word (e.g. Ross, Saint-Amour, Leavitt, Javitt, & Foxe 2007) in a single type of noise condition (e.g. Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009; Ross et al., 2007; Schwartz et al., 2004). Neglecting to simulate conditions present in daily communicative environments limits our understanding of the full scope of an individual's speech perception-in-noise ability.

An array of subgroups have been identified with speech perception-in-noise deficiencies, which provides an additional impetus to better understand the impact of adverse listening conditions on speech perception. These individuals range from those with neurodevelopmental disabilities, such as autism spectrum disorder (e.g. Alcantara, Weisblatt, Moore, & Bolton, 2004; Bishop & McArthur, 2005), sensorineural hearing loss (e.g. Helfer & Wilber, 1990), as well as individuals communicating in their non-native language (e.g. Mayo, Florentine, & Buus, 1997; Van Engen & Bradlow, 2007; Van Engen, 2010).

An estimated 20% of the U.S. population speaks a language other than English (U.S. Census Bureau, 2009). Therefore, based on student enrollment figures, one can extrapolate that there are approximately 16 million students growing up in a bilingual environment in the United States. Previous studies have revealed a discrepancy between monolingual and bilingual performance on speech-in-noise tasks, revealing that both bilingual children and adults are outperformed by their monolingual peers (e.g. Mayo et al., 1997; Nelson, Kohnert, Sabur, & Shaw, 2005), indicating that these students may face even greater deficits from adverse listening conditions in the classroom. However, it should be noted that these studies have primarily investigated the performance of non-native listeners when the speech-in noise task is presented in the listener's second language (e.g. Mattys, Carroll, Li, & Chan, 2010), or have predominately recruited children and adults whose families immigrated to the United States and learned English as a second language (e.g. Crandell & Smaldino, 1996). Thus, there is paucity of evidence regarding the performance on speech-in-noise tasks by simultaneous bilingual children and adults performance with a high proficiency in both of their languages.

In conclusion, it is important to provide further evidence exploring the underlying auditory and visual modalities of speech perception-in-noise, and to specifically observe the level of increase in intelligibility of speech signals when visual cues are available. This knowledge will allow teachers, professionals, clinicians, and parents to better understand the developmental trajectory of audiovisual integration and the impact of adverse listening conditions on speech perception-in-noise, and its impact on learning in the classroom. In turn, this knowledge will facilitate future development of optimal listening conditions for child and adult students, and may also contribute to future training methodology to aid groups of students who find it especially challenging to work against classroom listening conditions.

DEVELOPMENT OF SPEECH PERCEPTION

Speech perception requires the modulation of the peripheral and central auditory systems, coupled with the activation of cognitive abilities, such as attention and inhibition, in order to make sense of ambient speech signals. This complex task involves not only sensory processing, but also cognitive processing at higher cortical structures (Kraus & Chandrasekaran, 2010), where the ability to discriminate relevant information and decode meaning from the speech signal occurs. The human peripheral auditory system is advanced in anatomical development - many aspects of basic auditory processing appear to be adult-like within the first six months of an infant's life (Werner, 2007). These prolific structures enable early speech perception, which is an integral component of the language acquisition process, as it allows for the initial perception and processing of spoken language (Dawes & Bishop, 2009). Some contend that although infants enter the world prepared to perceive the ambient sounds around them, the complex central auditory processes, which are responsible for more advanced auditory processing, such as sound source segregation, require longer time to fully develop (Eggermont, 1985; Werner, 2007).

It is well established that the peripheral auditory system develops relatively early in life (Eggermont, 1985; Werner, 2007), however, there is still much left unknown about the protracted development of the complex central auditory processes. These processes have been demonstrated to continue to develop throughout at least the first decade of life (Ponton, Eggermont, Kwong, & Don, 2000). Behavioral tasks such as word recognition in noise (Elliot, 1979; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000), masking level difference (Hall, Buss, Grose, & Dev, 2004), and auditory sound source segregation (Sussman, Wong, Horvath, Winkler, & Wang, 2007) have been utilized to investigate the developmental trajectory of the central auditory process and the role it plays in speech perception.

Not only have behavioral tasks been utilized to demonstrate the increase of complex auditory system proficiency throughout childhood, but in some studies, these tasks reflect development to continue through adolescence into adulthood (Hazan & Barrett, 2000; Stuart, 2005). For example, Hazan & Barrett (2000) investigated the development of phonemic categorization and found that phonemic identification increased significantly between the ages of six and 12. Interestingly, the findings of this study revealed that, even at age 12, children were unable to categorize the phonemic contrasts as consistently as adults. Speech perception studies that have compared both child and adult participant performance have also demonstrated that the interference of auditory noise is a greater distractor in child participants (Barutchu et al., 2010; Riley & McGregor, 2012). It has further been indicated that the ability to detect speech in noise increases between 5 years of age and early adolescence (Johnson, 2000). However, this is a large age range and due to different experimental procedures used across studies (e.g. picture-word vs. speech-in-noise task), the question remains whether or not performance reflects developmental stage differences or the result of different task demands (Barutchu et al., 2010; Jerger et al., 2009).

THEORETICAL PERSPECTIVES: THE DEVELOPMENT OF AUDIOVISUAL INTEGRATION

There is a clear theoretical divide that has emerged with the goal to describe the development of audiovisual integration. For the purpose of this paper, we define audiovisual integration as the fusion of auditory cues (i.e. speech signal) and visual cues (i.e. articulatory facial movements) in order to form coherent representations of the environment (Barutchu et al., 2010). The divide predominately falls into two perspectives: 1) audiovisual integration is present early in an infant's life (e.g. Alridge, Braga, Walton, & Bower, 1999; Bahrick, Hernandez-Reif, & Flom, 2005; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) and 2) audiovisual integration develops over time through learning and experience (Ross et al., 2011; Sowell et al., 2004; Jansen, Chaparro, Downs, Palmer, & Keebler, 2013; Jerger et al., 2009). To date, there has been more conclusive evidence to support the latter hypothesis. However, the trajectory of AV development remains unclear, as there is a dearth of evidence of reflecting the integration of these processes in school-age children, with only a few behavioral and neural studies to date (e.g. Barutchu et al., 2010; Brandwein et al., 2011; Jerger et al., 2009; Moore, 2002).

The ambiguity of the developmental trajectory of audiovisual integration has led to the advancement of not only behavioral studies, but also neural studies (e.g. electrophysiological methods and functional neuroimaging). The majority of evidence supporting the notion that audiovisual integration is present early in life is found through both behavioral and neurological studies on infants as young as 2 months old. For example, Patterson & Werker (2003) used isolated vowels to demonstrate an early connection of auditory and visual systems in speech and found that infants as young as 2 months old had the ability to match phonetic vowel information to the correct articulation via facial presentation. Contrary to the evidence that has been provided for infants, audiovisual modalities investigated in school-age children demonstrate that visual articulatory speech cues have less impact on speech perception (Jerger et al., 2009).

Fortunately, a recent progression of neural studies has shed light on the neurophysiological changes that occur with the maturation of audiovisual multimodal functionality. Sowell et al. (2004) found evidence for the brain's audiovisual developmental trajectory by observing the cortical anatomy in perisylvian language areas. The authors revealed that this particular cortical area undergoes a relatively long developmental trajectory, supporting the theory that the fusion of the auditory and visual systems develop over time. In contrast, evidence has demonstrated that the cortical regions fundamental to basic sensory and perceptual functions develop before the perisylvian regions (Shaw et al., 2008). However, Ross et al. (2011) posit that the neural structures underlying audiovisual integration in speech develop concurrently with the higher-level language processes throughout adolescence.

Jansen et al. (2013) further expounded upon the initial findings of Sowell et al. (2004) and suggested that fully developed audiovisual integration depends on a combination of vision, audition and cognition. Results of their study reveal that for the typically developing adult, these modalities are fully developed. In contrast, in observing typically developing children, although visual and auditory modalities are present, their brain is still undergoing development and, therefore, the fusion of modalities is incomplete. This provides evidence demonstrating that neural connections between auditory and visual pathways for speech follow a developmental trajectory. With individual diversity observed across age groups, and the complexity of central auditory processes, it is all the more important to continue behavioral studies in order to guide and supplement neural studies and vice versa.

ADVERSE LISTENING CONDITIONS IMPACT ON SPEECH PERCEPTION

Mattys, Davis, Bradlow and Scott (2012) define adverse listening conditions as any suboptimal factor that may lead to a decrease in speech intelligibility on a given task, when performance on that same task is compared to the individual's performance in an optimal listening condition. The possible adverse listening condition factors are described as both external (i.e. the speaker and the speaking manner, the listener, and environmental noise), as well as internal (i.e. cognitive demands and compensatory strategies). It is well established that the intelligibility of speech perception-in-noise is modulated by the specific type of background noise or masker in which the speech signals are presented (Cooke, Lecumberri & Barker, 2008).

Energetic and informational maskers have been found to differentially modulate audiovisual speech integration in both adults and children. For example, one observed difference among maskers has been demonstrated through the notion of *glimpsing*, which describes the spectrotemporal regions at which a target signal is least impacted by the masker and, in turn, provides some amount of phonetic information (Cooke, 2006). To date, evidence indicates that children demonstrate lower accuracy on speech-in-noise tasks requiring the identification of final words in sentences presented in multiple-talker babble when compared to older peers and adults (Elliot, 1979; Fallon, Trenhub, & Schneider, 2000). The lower accuracy performance by younger school-age children has also been demonstrated when words and sentences are presented in spectral noise (Nittrouer & Boothryd, 1990).

Helfer and Freyman (2005) specifically investigated the interaction between visual information and the masking environment in adult participants. The experiment tested sentence intelligibility in the presence of steady-state noise and a two-talker masker, revealing that visual information was most salient to speech intelligibility in the presence of the speech masker as opposed to the steady-state noise. The authors posit that visual articulatory cues supplement the recovery of masked phonetic information as well as assist the listener in segregating the target from competing speech. Therefore, based on this evidence, employing multiple types of maskers to standard speech-in-noise batteries will lead to further insight into audiovisual integration and the enhancement of intelligibility due to observed visual cues. However, before looking at speech-in-noise tasks across age groups, one must first understand the divergent theoretical perspective regarding audiovisual integration in adverse listening conditions.

THEORETICAL PERSPECTIVES: AV INTEGRATION IN SPEECH PERCEPTION-IN-NOISE

There are two predominant and competing hypotheses that have been presented to explain audiovisual integration in speech perception in noisy environments. The first is the principle of inverse effectiveness (PoIE), which Meredith & Stein (1986) derived to explain audiovisual integration in speech perception. According to this principle, audiovisual integration benefits speech intelligibility the most when the signal-to-noise ratio (SNR) between auditory speech signals and interfering noise levels is most difficult (Sumby & Pollack, 1954; Eber, 1969; Eber, 1979).

In contrast to Meredith & Stein, Ross et al. (2007) found evidence to support a window of maximal multisensory integration beyond the predictions of the PoIE at the intermediate signal-to-noise ratio (SNR) of -12 dB. Ross et al. used a range of SNRs (0 to

-24 dB) to examine speech perception-in-noise. The findings of this study indicate that that maximal audiovisual integration occurred at -12 dB, rather than the most difficult SNR condition (i.e. -24 dB) (Ross et al., 2007; Ross et al., 2011; Ma et al., 2009).

However, this interaction may not be so easily explained through a single hypothesis. For example, recall that different maskers modulate speech perception in noise differentially, and therefore influence the degree to which visual cues are utilized. Recent studies suggest that the audiovisual integration in speech perception in noise may primarily depend upon the type of background masker (Helfer & Freyman, 2005; Bernstein & Grant, 2009). For example, in the Helfer & Freyman study, the visual gain in speech intelligibility was approximately 5.5 dB larger for informational masking when compared to performance in energetic masking. Moreover, visual cue benefit was found to differ qualitatively across the two masking conditions. That is, in energetic masking, visual cues are utilized more at an intermediate level of SNR (-12dB) (e.g. Ross et al., 2007), while in informational masking, when both the masker and the signal are speech stimuli, the perception of the spatial separation between the speech signal and the masker can be adequate for a significant speech recognition advantage to occur (Arbogast, Mason, & Kidd, 2002). Thus, the benefit of visual cues may be less susceptible depending on the masker type.

SPEECH PERCEPTION-IN-NOISE AND THE DEVELOPMENT OF AV INTEGRATION

Previous studies have demonstrated that the ability to perceive unimodal auditory speech when it is masked in noise develops with age (Barutchu et al., 2010; Hetu et al., 1990; Johnson, 2000). Emerging evidence has indicated similar developmental results for multimodal audiovisual speech perception-in-noise. As aforementioned, one explanation for the development of multisensory speech perception is from the neurological perspective: as we age, the auditory and visual areas of the brain mature to provide us with a reliable source of perceptual information (McLeod, 2007). To support this hypothesis, Ross et al. (2011) conducted an audiovisual speech-in-noise experiment to investigate the pattern found in previous imaging studies, which indicated that the perisylvian cortex (a neural correlate associated with speech and language functions) continues to develop later into childhood. The authors measured word recognition in children (age range=5-14) and adults by presenting audiovisual stimuli at various levels of SNR. The findings validate the imaging studies, and further demonstrate that the integration of audiovisual cues in speech perception-in-noise tasks improve accuracy more in adult participants.

To investigate the behavioral findings of Ross et al. (2011), Knowland, Mercure, Karmiloff-Smith, Dick, and Thomas (2014) observed the utilization of visual speech cues in speech perception-in-noise tasks combined with an event-related potential (ERP) task, comparing children (age range=6-11) to adults (age range=20-34). They found that audiovisual modalities undergo a gradual maturation over mid-to-late childhood. The authors conclude that visual speech is represented by separate underlying cognitive processes that mature earlier compared to other cognitive processes at different stages of development.

One explanation for the observed difference in adult and child performance is the child’s limited language experience, and to that end, some studies have compared child participants to adult native speakers of English. For example, native speakers are more proficient at identifying speech-in-noise than are non-native speakers with several years of exposure to English (Mayo et al., 1997; Van Engen, 2010; Van Engen & Bradlow, 2007). This could be due to the fact that throughout the lifespan, as words become increasingly familiar, less acoustic information is required for their identification (Van Engen, 2010). Therefore, from the current research it can be assumed that the visual benefit, or the window of maximal visual benefit pattern at -12 dB, must also emerge during childhood as auditory, visual, and cognitive systems develop (Ross et al., 2007).

CULTURAL FACTORS TO CONSIDER IN SPEECH PERCEPTION: BILINGUALISM

The term bilingualism is not easily defined. Baker (1993) defined the term bilingual as an individual who knows two languages. However, with the progression of bilingual research, this definition will not suffice. Throughout the literature, bilinguals are now defined broadly by their early or late onset of a second language, or more stringently simultaneous or sequential (for a review, see McLaughlin, 2013). Over the past decade, with an increase in new findings, a better understanding of the external and internal factors that are found within Baker's broad definition have emerged, demonstrating that this heterogeneous group differs in age of acquisition, level of proficiency and amount of language usage (Paradis, 2011).

At the early stages of bilingual research, many professionals believed that bilingualism negatively impacted cognitive and linguistic development, inhibiting full intellectual potential in typically developing individuals (for reviews, see Cummins, 1976; Diaz, 1983). However, according to Bialystok (2010), research over the past several decades has disproven this initial hypothesis, and in turn, has provided evidence for possible cognitive strengths, such as inhibition and executive control, in typically developing bilingual individuals when compared to their monolingual age-matched peers.

Therefore, the literature concludes that bilingualism either elicits a positive effect in linguistic domains, e.g. enhancing metalinguistic awareness, or no effect on intelligence at all (Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, 2010). Current bilingual research has further corroborated cognitive strengths in typical bilingual individuals, and has revealed executive control, problem solving, creativity as well as inhibitory strengths in bilingual individuals when compared to monolingual peers (e.g. Bialystok & Martin, 2004; Blumenfeld & Marian, 2011; Goetz, 2003).

Over the past several decades, researchers have sought to better understand the peculiarities of bilingual language processing. The impetus for this body of interdisciplinary research stems from the fact that bilinguals constantly face a higher cognitive demand, compared to monolingual peers. For example, bilingual individuals are able to switch between two languages without letting the lexicon of their inactive language seep into their activated spoken language (for reviews, see Marian, 2009; Kroll, Gullifer, & Rossi, 2013). There is much debate as to the exact manner and method that bilingual individuals employ in order to match linguistic input to one of their languages.

Dijkstra (2005) highlighted two deviating hypotheses that have sought to better define and capture the bilingual language selection process. The first is described as the *language-selective access hypothesis*, which indicates that bilinguals possess two independent lexical systems that are selectively accessed, depending upon linguistic input. This hypothesis indicates that the two languages of the bilingual are stored and processed separately, and when one language is used the bilingual mind then behaves like a monolingual in selecting and using only one language (Kroll et al., 2013). Contrary to this hypothesis, the *nonselective access hypothesis* posits that bilinguals possess an integrated lexicon, in which, during word recognition and selection process, lexical representations from both languages are simultaneously activated. Evidence from neuroimaging studies has proven the latter, supporting the notion that a co-activation of linguistic knowledge, rather than an individual selection of both languages occurs when bilinguals read, speak, and listen to speech in one language alone (Bialystok & Martin, 2004; Bialystok, 2010; Dijkstra, 2005).

BILINGUAL SPEECH-IN-NOISE PERFORMANCE COMPARED TO MONOLINGUAL PEERS

There is significant evidence that demonstrates that early bilinguals appear to have an advantage over monolinguals in the cognitive domain in the areas of problem solving and creativity (Bialystok, 2010; Kessler & Quinn, 1987), as well as executive function, memory, cognitive inhibition, and attention (Bialystok et al., 2004; Bialystok & Martin, 2004; Blumenfeld & Marian, 2011). The greater cognitive demands placed on bilingual language processing has been a fundamental explanation for the bilingual advantage. Greater cognitive demand has been demonstrated in the bilingual speaker's ability to switch between two different languages (i.e. code-switching), and also has been explained through the individual's ability to suppress a second language during speech production (Dikstra, 2005). An array of interdisciplinary experiments have been developed to investigate the bilingual advantage hypothesis, spanning from electroencephalography, functional magnetic response imaging, and eye-tracking tasks, to non-linguistic behavioral based tasks such as the Stroop task. For example, Blumenfeld and Marian (2011) utilized an eye-tacking/negative priming task and collected information on both the activation of multiple word candidates during auditory comprehension and subsequent suppression of irrelevant competing words. The authors demonstrated that inhibitory performance on a nonlinguistic Stroop task was related to linguistic competition resolution in bilinguals, but not in monolingual age-matched peers. Speech perception-in-noise tasks have also been identified as useful tools in order to further explore these posited bilingual advantages, as one would hypothesize that the greater inhibitory control found in bilinguals may result in their better separation of the target speech signal from noise, when compared to monolingual peers (Marian, 2009).

There is significant evidence that has revealed that both early and late bilinguals demonstrate *lower* performance in speech perception tasks under adverse listening conditions compared to monolingual listeners (e.g. Mayo et al., 1997; Bradlow & Bent, 2002; Cutler et al., 2004; Rogers et al., 2006; Von Hapsburgh & Bahng, 2006; Bovo & Callegari, 2009; Tabri, Chacra, & Pring, 2011). Previous studies have specifically demonstrated that, although monolingual and bilingual listeners perform similar in quiet conditions, bilingual listeners require an easier SNR (on average, about 8 dB) in order to perform similarly to monolingual peers in adverse listening conditions (Van Engen, 2010). However, to date no studies have examined bilingual performance using audiovisual speech perception-in-noise conditions. Those that have explored audiovisual integration in bilinguals have utilized nonlinguistic tasks to reflect attention and inhibition abilities (e.g. Stroop task) and have hypothesized that these evidenced strengths in bilinguals would generalize to greater audiovisual processing in proficient bilinguals when compared to monolingual peers (Marian, 2009).

One predominant factor that makes it difficult to converge on a conclusion regarding bilinguals performance on speech perception-in-noise tasks in due to the fact that all of the studies do not define bilingualism in the same manner, and the majority of past research was conducted on non-native listeners who were described as late bilinguals acquiring English after age 6 (e.g. Mayo et al., 1997; Rogers et al., 2006). Attempting to remediate the paucity of evidence for early bilinguals with high proficiency in the English language, Rogers et al. (2006) sought to investigate speech in noise task performance in adults defined as "early bilinguals", those who have acquired a second language before age 6. The recruited participants were highly proficient Spanish-English bilinguals who were reported to have no accent in English. The results on a monosyllabic word recognition task in speech-shaped noise and reverberation conditions revealed that although monolingual and bilingual performance was comparable in quiet conditions, monolingual participants' accuracy exceeded bilingual age-matched peers' as SNR became more difficult.

Rogers et al. (2006) and Blumenfeld and Marian (2011) proposed competing hypotheses in regard to bilingual performance on speech-in-noise tasks. According to Rogers et al. (2006), bilingual listeners are disadvantaged on speech-in-noise tasks as a result of increased demand for attentional resources and increased processing demand. Rogers et al. (2006) further posit that this may be due to the bilinguals' need to deactivate the inactive language, to select target phonemes from a larger number of alternatives, or to match native speaker productions to phonetic categories that may be between the norms for their two languages. It would be remiss not to recognize that, although this line of research supports the hypothesis of the language-access-selective hypothesis, there are still observed bilingual advantages in inhibitory and controlled processing, as observed in the study conducted by Blumenfeld and Marian (2011). Therefore, in observing the findings of these researchers, one may still predict a bilingual advantage for speech perception in speech-in-noise tasks in highly proficient simultaneous bilingual speakers. That is, speech-in-noise requires cognitively suppressing irrelevant information during co-activation of both languages, while focusing on target information, an ability that appears to be enhanced in bilinguals through the nonlinguistic Stroop task.

RATIONALE FOR THE CURRENT STUDY

A review of the literature indicates that visual cues can significantly enhance a degraded auditory speech signal to improve intelligibility to a degree equivalent to increasing the signal-to-noise ratio by 15 dB (e.g. Sumby & Pollack, 1954). However, there is a paucity of evidence demonstrating this increased intelligibility in school-age children. Moreover, there is a dearth in evidence providing information for both school-age and university-age simultaneous bilingual students with high proficiency and usage of both languages. Ross et al. (2011) demonstrated that visual speech information can improve the comprehension of speech recognition, and additionally confirmed the developmental trajectory of audiovisual modulation in speech perception-in-noise by comparing both child and adult participants. However, the authors only presented words in one type of masker (i.e., energetic). In the typical classroom environment, noises are presented not only in the form of a loud heating and cooling units, but also in the form of other children chatting in the back of the room, in the hallway adjacent to the classroom door, or yelling outside the window on the playground. Therefore, without the implementation of informational maskers in speech perception-in-noise experiments there remains a gap in knowledge identifying when and how the auditory and visual systems come to work together in development and how these modalities are impacted by different types of everyday adverse classroom listening conditions.

Further research is needed in order to increase our understanding of the developmental trajectory of audiovisual speech perception, as well as the way bilingualism modulates audiovisual integration during speech perception-in-noise tasks. The aim of this project is two-fold: 1) to investigate developmental differences in intelligibility gains from visual cues in speech perception-in-noise, and 2) to examine how different maskers modulate visual enhancement across age groups.

A secondary aim of this project is to investigate the extent to which bilingual experience differentially modulates audiovisual processing. This investigation will contribute to our understanding of the multimodality of language processing in bilinguals, and provide further insight into the specific advantages and disadvantages regarding speech perception-in-noise for this population. We seek to specifically determine if a more diverse linguistic input across multiple modalities in bilingual speakers generalizes to a greater utilization of visual cues.

In conclusion, the current study investigates the impact of maskers on speech intelligibility across various age groups on speech perception-in-noise tasks. We predict that bilingual speakers, both children and adults, will rely more on visual cues as listening environments become increasingly difficult. This is because bilingual speakers have a more diverse linguistic input and therefore are expected to rely more on multimodal integration in speech perception. Our study is one of the first to investigate the impact of bilingualism on audiovisual processing and speech perception-in-noise, in both school-age and adult students.

METHODOLOGY

CHILD PARTICIPANTS

Thirty children (14 monolingual and 16 bilingual speakers, age range=6-10, mean age=7.4) were recruited from Great Wall China Sunday School and St. Elias Orthodox Church School. The first language for all participants was English. Each child was born in the United States and did not spend any time outside the country. The 14 monolingual speakers (6 females; 8 males; age rage=6-10; mean age=7.6) parents reported that their child did not have significant exposure to a second language throughout their lifespan. The 16 bilingual speakers (8 females; 8 males) consisted of 8 English-Chinese, 4 English-Arabic, 3 English-Swedish participants, and 1 English-Spanish participant. All parents of bilingual participants reported that their child's daily use of second language exceeded 20%. All participants were current elementary students in Austin, TX. Each participant completed a pure tone hearing screening (sweep test) to ensure thresholds of <20 dB HL at 1000 Hz, 2000 Hz, and 4000 Hz. All child participants, as well as their parents, provided written informed consent. Parents of both monolingual and bilingual participants completed respective background forms. The general nonverbal intelligence of each child participant was assessed using the Kaufman Brief Intelligence Test, Second Edition (KBIT-2). An analysis of variance (ANOVA) revealed that monolingual and bilingual child participants did not differ in intelligence or socioeconomic status. Upon completion of all experiment procedures, children received $10 compensation as well as a prize for their participation.

ADULT PARTICIPANTS

Thirty-one adults (age range=18-27, mean age=20.5) were recruited from the University of Texas at Austin. The first language for all participants was English. Each adult was born in the United States and did not spend significant time outside the country. The 21 adult monolingual speakers (10 males; 11 females; age range=18-27; mean age=20.9) all spoke English as their first language and reported that they did not have significant exposure to a second language until high school to meet foreign language curriculum requirements.

The 10 adult bilingual speakers (2 males; 8 females) consisted of 4 English-Spanish, 3 English-Chinese, 2 English-Korean, and 1 English-Urdu participant. All bilingual adult participants were categorized as simultaneous bilinguals, indicating that they were exposed to both English and their second language simultaneously from birth.

Every adult participant was either a current undergraduate or graduate student. Each participant completed and passed a pure tone hearing screening (sweep test) to ensure thresholds of <20 dB HL at 1000 Hz, 2000 Hz, and 4000 Hz. All adult participants provided written informed consent. Both monolingual and bilingual adult participants completed respective background forms, to control for second language onset, daily language usage, socioeconomic status, and presence of a developmental disability. The general nonverbal intelligence of each adult participant was assessed via the Kaufman Brief Intelligence Test, Second Edition (KBIT-2). Upon completion of the experiment adult participants were compensated $10 for their participation.

Both child and adult bilinguals were considered to be simultaneous bilinguals based on subgrouping methodology by McLaughlin (2013), who used a cutoff of 3 years, based on the fact that this is the age that typical developing children have phrase-level expressive language abilities.

	Child Participants		Adult Participants	
	Monolingual	Bilingual	Monolingual	Bilingual
N	14	16	21	10
Age	7.6 (1.3)	7.2 (1.1)	20.8 (2.1)	19.9 (1.5)
SES-mother	46.6 (15.9)	37.2 (21.1)	39.0 (14.8)	33.0 (15.5)
SES-father	53.1 (16.8)	61.6 (6.4)	49.4 (19.4)	53.8 (15.6
SES-family	57.0 (6.5)	55.2 (12.4)	50.8 (12.1)	52.8 (14.9)
KBIT-standard	107 (18.5)	110 (22.3)	106 (11.0)	109 (10.7)
L1 % daily use		54.7 (29.4)		76.9 (10.5)
L2 % daily use		45.3 (29.4)		21.5 (8.8)
L1 age of acquisition		1.28125		0
L2 age of acquisition		0		0

Table 1. Analysis of Variance for Participant Descriptive Data

TEST MATERIALS

All experiments and procedures for this study were approved by the Institutional Review Board of The University of Texas at Austin.

Background Questionnaires

Monolingual General Background Questionnaire

Additional demographic information was collected from the monolingual adult participants and the child participants via parents, in order to control for socioeconomic status, hearing ability, and the presence of a developmental disability.

The Language History Questionnaire (LHQ 2.0)

The LHQ 2.0 (Li, Zhang, Tsai, & Puls, 2013) is a web-based tool for collecting linguistic background information from bilinguals or second language learners, and is a proven methodology for analyzing the self-reported proficiency of bilinguals. The authors based their questionnaire on the most commonly asked bilingual questions across published studies (for full description see Li et al., 2013). Adult bilingual participants completed the web-based LHQ 2.0, which provided them with a private means for completing the questionnaire, since their identity was protected through the assignment of a unique ID number.

Parent Bilingual History Questionnaire

Empirical evidence indicates that parents of bilingual children are reliable reporters of language development (Dale, 1991). Therefore, information about the bilingual children's language use and proficiency level was collected through a parent bilingual history questionnaire (as described in Sheng, Lu, & Kan, 2011), as well as through an informal parent interview. The family history and speech-language development sections of the original parent bilingual history questionnaire were modified in order to better correlate with questions from the adult LHQ 2.0. Parents were asked about the people with whom the child interacted in different settings (school vs. home), on different days of the week (weekdays vs. weekend), as well as the child's preferred language of communication across settings (second language, English, or both).

Yale Journal of Sociology Four Factor Index of Social Status

The Yale Journal of Sociology Four Factor Index of Social Status was utilized to calculate reliable socioeconomic scores for each participant and control for socioeconomic environment. The Social Striatum for each participant was derived by a four factor index of social status which equals: occupation × education × gender × marital status. All participants' family Social Striatum in this study fell into two categories: 1) medium business, minor professional, technical (Social Striatum range=54-40) or 2) major business and professional (Social Striatum range=66-55). An analysis of variance revealed no significance difference among participants, both children and adults.

Kaufman Brief Intelligence Test-Second Edition (KBIT-2)

The nonverbal matrices subtest of the KBIT-2 was administered to assess the nonverbal intelligence for all participants (Kaufman & Kaufman, 2004). This assessment tool has been normed for age range=4:0-90:0, and therefore could be administered to all participants. This particular subtest consists of 46 items divided into three sections of increasing difficulty. On each trial, the child or adult was presented with visual stimuli representing either drawings of concrete objects or abstract figures. The first portion of the test consisted of one target at the center of the page and five potential picture answers below the target, while the latter portion of the assessment prompted the child or adult to complete an incomplete display of 2 × 2 or 3 × 3 matrices. The standard procedure as described in the administrator’s manual was utilized for testing and scoring.

EXPERIMENT MATERIALS

Target Speech Sentences

One male native speaker of American English was video-recorded producing one set of sentences on a sound attenuated stage at The University of Texas at Austin. 80 semantically meaningful sentences were recorded based on sentences from the Basic English Lexicon (BEL) (Calandruccio & Smiljanic, 2012). Sentences consisted of 4 keywords each (e.g. The HOT SUN WARMED the GROUND; see appendix). All sentences were produced in a conversational speaking style. To elicit this speaking style, the speaker was prompted to speak as if he were talking to a familiar listener. A Sony

PMW-EX3 studio camera was used as the video recorder for the target sentences, and enabled each sentence to be presented to the speaker via teleprompter. Camera output was processed through a Ross crosspoint video switcher and recorded on an AJA Pro video recorder. Audio was recorded at a sampling rate of 48000 Hz with an Audio Technica AT835b shotgun microphone placed on a floor stand in front of the speaker. One long initial video recording of the speaker producing all 80 sentences was completed, followed by the segmentation of each individual sentence. Following this procedure, Final Cut Pro software was utilized to extract the audio from each sentence video file. Praat software (Boersma et al., 2009) was then utilized to equalize the RMS amplitude. The leveled audio clips then became the auditory stimuli for the audio-only (AO) condition. The leveled audio files were then reattached to the corresponding video files using Final Cut Pro. Stimuli consisted of 80 sentences with 4 target words each. All sentences were produced by the same native English male speaker.

Maskers

Each sentence was masked by one of two types of noise: 1) informational masking: a 10 second masker track of two-talker babble (2T); 2) energetic masking: a 10 second masker track of pink noise (P). The two-talker babble track was created by two male native, American English speakers recorded in a sound-attenuated booth at Northwestern University as part of the Wildcat Corpus project (Van Engen et al., 2010). Each participant produced a set of 30 simple, meaningful English sentences (Bradlow & Alexander, 2007). Each sentence was segmented from the recording files and equalized for RMS amplitude. The sentences from each talker were concatenated to create two tracks of 30-sentence strings with no silence between sentences. Next, these two tracks were mixed to generate a two-talker babble track. The final babble track was trimmed to 50 seconds.

The pink noise track and final babble tracks were both equated for RMS amplitude to 50, 54, 58, 62, and 66 dB SPL using Praat (Boersma et al., 2009) to create 80 noise clips. For each target sentence, there were five pink noise clips with increasing sound levels in the step of 4 dB SPL, and five two-talker babble clips with increasing sound levels in the step of 4 dB SPL. Each noise clip was 1 second longer in duration than its accompanying target sentence.

Mixing targets and maskers

All target sentences were segmented from the original long video recording. The audio was detached from each segmented video and RMS amplitude equalized to 50 dB SPL using Praat (Boersma et al., 2009). Each audio clip was mixed with 5 corresponding pink noise clips and 5 corresponding two-talker babble clips to create 5 stimuli of the same target sentence for each masker type with following SNRs: 0 dB, -4 dB, -8 dB, -12 dB, & -16 dB. The mixed audio clips then became the stimuli for the audio-only condition. The mixed audio clips were reattached to the corresponding video files to create the stimuli for the audiovisual condition. A freeze frame of the speaker was captured and displayed during the 500 ms noise leader and 500 ms noise trailer. In total, there were 400 final audio files and 400 corresponding audiovisual files with pink noise masker (80 sentences × 5 SNRs), as well as 400 final audio files and 400 corresponding audiovisual files with the two-talker babble masker (80 sentences × 5 SNRs).

PROCEDURES

Before the speech-in-noise experiment was administered, the participants signed an informed consent document and completed a pure tone sweep test following experiment protocol. In compliance with the American Speech-Language Hearing Association guidelines for manual pure-tone threshold audiometry, two positive elicited responses were recorded for frequencies at 1000, 2000, and 4000 Hz for each participant. Screening levels for all participants were at 20 dB, since all participants were over age 4 (which is the cut-off for sweep test at 25 dB). Controlled instructions were given to each participant to prepare for the screening. Experiment protocol instructed testing to be discontinued if two negative responses were elicited at any frequency. The experiment then took place in a sound-attenuated booth using E-Prime 2.0 software (Schneider et al., 2002). The sound stimuli were bilaterally presented to participants through Sennheiser headphones at a fixed 26 volume level.

There were three within subject variables: (1) masker type: pink noise or two-talker babble, (2) modality: audio-only (AO) or audiovisual (AV), and (3) SNR: 0 dB, -4 dB, -8 dB, -12 dB, and -16 dB. Each participant listened to four target sentences in each condition. There were 80 total trials for each condition. The 80 trials were mixed and presented to the participants in a randomized order. Therefore, the assignment of each sentence to a particular condition was randomized for each participant and no target sentence was presented more than once.

For child participants, the experiment was presented as a game in which they were encouraged to attend to the speaker that was presented on the screen, as well as the speech they were hearing through the headphones. The development of a game-like procedure for child participants was motivated by past child studies that indicate the importance of attention maintenance in child subjects to ensure optimal test performance (Dawes & Bishop, 2008). Game instructions were directly read from the screen to each child participant. Their task was to listen carefully and to make their best guess regarding what the speaker just said. "For this game, you will listen to 80 sentences mixed with different types of noise. The noise might sound like static on a television or a bunch of people talking in a restaurant. Sentences will either be presented with the sound only, or they will also have a video of the speaker."

One trained research assistant was present to type the child's percepts and ensure that the child was paying attention to the screen and speaker presentation at all times during the experiment. The child was instructed that the objective of the game was to first listen to the sentence the speaker says, and then repeat the exact sentence that they heard out loud. The child was further instructed that the speaker would begin talking after the noise. Finally, the child was instructed that even if they only heard a few words, to say those words out loud, and if they were unsure to make their best guess. If they did not understand any words, they were asked to say 'X'.

The only difference between the child and adult experiments was that in the adult experiment each trial was self-initiated by the adult by pressing a key on a keyboard. The adults were instructed to type the target sentence after stimulus presentation. If they were unable to understand the entire target sentence, like the child participants, they were prompted to make their best guess and report any intelligible words heard. If they did not understand any words, they were asked to type 'X'.

For trials in the audio-only condition, a centered black cross on a white background was presented on the screen concurrently with the sound stimulus; for trials in the audiovisual condition, a full-screen video of the speaker was presented along with the sound stimulus. Before the experiment, adult participants were instructed that they would listen to sentences mixed with noise and that each sentence would either be audio-only or accompanied by a video of the speaker. They were also informed that the target sentences would always begin one-half second after noise onset.

DATA ANALYSIS

Speech Intelligibility Accuracy: Participant reported responses were scored per accurately typed keyword. Responses that included homophones and phonetic misspellings were scored as correct. The proportion of correctly identified keywords was then calculated for each experimental condition for all participants. The intelligibility data was analyzed with a linear mixed effects logistic regression (LMER) where keyword identification (correct vs. incorrect) was the dichotomous dependent variable. Subjects were included in the model as random factors, and SNR, modality, listener group, and their interactions as fixed effects. SNR was mean-centered as a continuous variable. Modality and listener group were treated as categorical variables. Analysis was performed using the lme4 package in R (Bates, Maechler, & Bolker, 2012).

Visual enhancement: At each SNR, visual enhancement (VE) was calculated as the performance difference between the AV and AO condition, using the formula: VE=AV-AO (Ross et al., 2007). This index quantified the AV processing benefit to speech intelligibility at each SNR.

RESULTS

Adopting a developmental perspective, our subsequent analyses focus on comparing children's ability to process speech-in-noise to that of adults in the presence or absence of visual cues. In addition, we examined the possible effect of bilingualism on such ability. Participants' performance, operationally defined by correct keyword identification, was analyzed with a linear mixed effects logistic regression (LMER) wherein keyword identification (correct or incorrect) was treated as a dichotomous dependent variable. Subjects were included in the model as random factors, while language group (monolingual vs. bilingual), age group (child vs. adult), SNR (0 dB, -4 dB, -8 dB, -12 dB, -16 dB), masker type (two-talker babble vs. pink noise), and their interactions were included as fixed effects. Language group, age group, and masker type were treated as categorical variables. SNR was mean-centered and treated as a continuous variable. Analysis was performed using the lme4 package in R (Bates et al., 2012).

AO condition Before examining the change in performance across SNR, we compare the overall performance in each masker condition. Analysis reveals a statistically significant age group × masker type interaction ($p<.001$) and age group × masker type interaction × language group interaction ($p=.04$). Further breakdown of the higher order 3-way interaction revealed that change in masker-type brings along opposite effects for children and adults (Table 2). While children performed better in the pink noise condition (mean accuracy correct=38%) than in two-talker condition (mean accuracy correct=32%) ($p<.001$; Table 3), adults performed better in the two-talker condition (mean accuracy correct=70%) than in the pink noise condition (mean accuracy correct=55%) ($p<.001$; Table 4). Figures 1 and 2 demonstrate this interaction. With regard to the incremental improvement across elevation of SNR, a 4-way language group × age group × masker type × SNR interaction was found and the lower order interaction was not analyzed. We examined this interaction by looking at the performance in pink noise (Table 5) and two-talker conditions separately (Table 6). In both conditions the effect of SNR is significant ($p<.001$), wherein elevation in SNR increased the probability of correct identification of keywords. In both conditions the age effect is significant ($p<.001$) and adults outperformed children. However, in the two-talker babble condition alone there is a significant 2-way age group × SNR interaction ($p<.001$) and a 3-way age group × language group × SNR interaction ($p<.001$). We further broke the higher order 3-way interaction down and found that it was driven by the difference between monolingual and bilingual children (Table 7) but not adults (Table 8). In the two-talker babble condition (2T), there is a statistically significant language group × SNR interaction in children ($p<.001$) but not in adult groups ($p=.28$). Here, the increase of SNR brings less improvement in monolingual children than in bilingual children (Fig. 2).

Fixed effects:	Estimate	Std. error	z value	p
(Intercept)	1.06	0.23	4.56	<.001
SNR	0.22	0.01	12.70	<.001
Masker type	-0.64	0.14	-4.51	<.001
Age group	-2.69	0.31	-8.60	< .001
Language group	0.25	0.29	0.89	.372
SNR:Masker type	0.24	0.03	7.47	<.001
SNR:Age group	0.19	0.02	6.72	<.001
Masker type:Age group	1.23	0.20	6.10	<.001
SNR:Language group	0.01	0.02	0.79	.426
Masker type:Language group	0.02	0.18	-0.11	.909
SNR:Masker type:Age group	-0.24	0.04	-5.49	<.001
SNR:Masker type:Language group	0.04	0.04	1.07	.280
SNR:Age group:Language group	-0.13	0.03	-3.70	<.001
Masker:Age group:Language group	-0.57	0.27	-2.10	.035
SNR:Masker:Age group:Language group	0.12	0.06	1.97	.047

Table 2. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AO condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	-0.81	0.12	-7.03	< .001
Masker type	0.32	0.08	3.75	< .001
Language group	0.06	0.17	0.38	.701
Masker type:Language group	-0.10	0.12	-0.81	.419

Table 3. Child Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AO condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	0.78	0.13	6.16	< .001
Masker type	-0.60	0.10	-5.93	< .001
Language group	0.16	0.16	0.99	.322
Masker type:Language group	-0.09	0.13	-0.68	.495

Table 4. Adult Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AO condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	0.41	0.20	2.02	.043
SNR	0.45	0.02	15.93	<.001
Age group	-1.41	0.26	-5.32	< .001
Language group	0.23	0.25	0.92	.354
SNR:Age group	-0.05	0.03	-1.41	.158
SNR:Language group	0.06	0.03	1.80	.071
Age group:Language group	-0.40	0.36	-1.11	.265
SNR:Age group:Language group	-0.01	0.05	-0.27	.786

Table 5. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in pink noise in AO condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	1.11	0.30	3.64	< .001
SNR	0.23	0.01	12.69	< .001
Age group	-2.82	0.40	-6.94	< .001
Language group	0.31	0.38	0.81	.417
SNR:Age group	-0.20	0.03	6.52	< .001
SNR:Language group	0.02	0.02	1.09	.275
Age group:Language group	0.14	0.54	0.25	.795
SNR:Age group:Language group	-0.15	0.03	-3.86	< .001

Table 6. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in two-talker babble in AO condition

Fixed Effects:	Estimate	Std. Error	z value	P
(Intercept)	-1.71	0.27	6.31	< .001
SNR	0.43	0.02	17.66	< .001
Language group	0.45	0.38	1.16	.245
SNR:Language group	-0.13	0.03	-4.02	< .001

Table 7. Child Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in two-talker AO condition

Fixed Effects:	Estimate	Std. Error	z value	P
(Intercept)	1.12	0.30	3.68	< .001
SNR	0.23	0.01	12.69	< .001
Language group	0.31	0.37	0.81	.413
SNR:Language group	0.02	0.02	1.08	.276

Table 8. Adult Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in two-talker AO condition

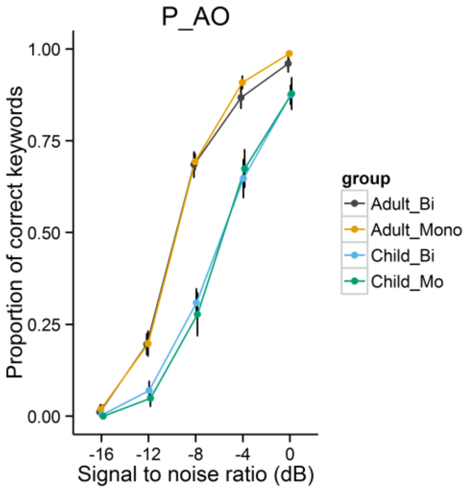

Figure 1. Performance in pink noise condition with audio-only

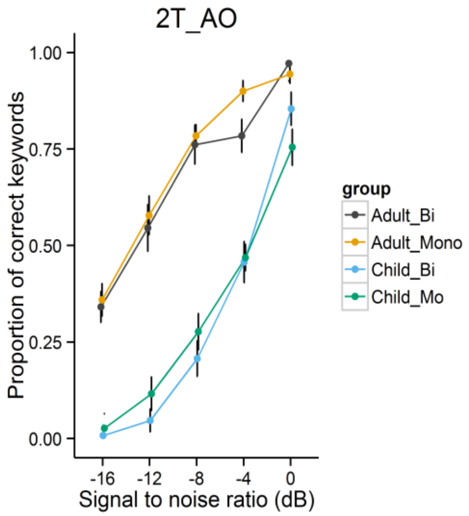

Figure 2. Performance in two-talker babble with audio-only

AV condition Audiovisual condition performance across all 5 SNRs was again collapsed in each masker condition respectively to examine the overall performance. Analysis reveals a significant age group × masker type interaction (p=.03; Table 9), wherein the child group's performance was higher in the pink noise (mean accuracy correct=48%) condition than in the two-talker babble condition (mean accuracy correct=44%; Table 10). In contrast, there was no statistical evidence to support the adult group performing differently across masker types (p=.92; Table 11). With regard to the incremental improvement across the increase in SNR, a two-way masker type × SNR interaction (p<.001) and a 3-way age group × masker type × SNR interaction (p<.001) was found and lower order interaction was not analyzed. In both two-talker babble (Table 12) and pink noise (Table 13) conditions there is a statistically significant SNR effect (p<.001) and age group effect (p<.001), but only in the two-talker babble condition is an age group × SNR interaction observed (p<.001), wherein increase in SNR brings a larger incremental improvement in the probability of correct keyword recognition in children than in adults. This suggests that the incremental improvement in performance is comparable between both age groups in pink noise but not in two-talker babble, which is likely because children perform more poorly in the latter condition (Figure 3; Figure 4).

Fixed effects:	Estimate	Std. error	z value	p
(Intercept)	1.52	0.23	6.59	<.001
SNR	0.14	0.01	8.71	<.001
Masker type	0.01	0.14	0.09	.921
Age group	-2.12	0.29	-7.12	< .001
Language group	0.57	0.29	1.96	.049
SNR:Masker type	0.13	0.02	5.00	<.001
SNR:Age group	0.15	0.02	6.71	<.001
Age group:Masker type	0.40	0.18	2.16	.030
SNR:Language group	0.01	0.02	0.75	.452
Masker type:Language group	-0.18	0.20	-0.92	.355
SNR:Masker type:Age group	-0.11	0.03	-3.20	<.001
SNR:Masker type:Language group	0.04	0.03	1.34	.178
SNR:Age group:Language group	-0.01	0.03	-0.41	.681
Masker type:Age group:Language group	-0.10	0.25	-0.40	.687
SNR:Masker type:Age group:Language group	-0.05	0.04	-1.14	.253

Table 9. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AV condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	-0.61	0.22	-2.74	.006
SNR	0.31	0.02	18.83	< .001
Masker type	0.42	0.11	3.86	< .001
Monolingual	0.37	0.32	1.16	.248
SNR:Masker type	0.02	0.02	0.95	.340
SNR:Language group	0.00	0.02	0.18	.856
Masker type:Language group	-0.29	0.16	-1.85	.064
SNR:Masker type:Language group	-0.01	0.03	-0.22	.827

Table 10. Child Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AV condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	1.52	0.19	8.21	< .001
SNR	0.15	0.02	8.70	< .001
Masker type	0.01	0.15	0.10	.923
Language group	0.56	0.24	2.38	.017
SNR:Masker type	0.14	0.03	4.99	< .001
SNR:Language group	0.02	0.02	0.74	.461
Masker type:Language group	-0.18	0.20	-0.91	.361
SNR:Masker type:Language group	0.05	0.04	1.32	.187

Table 11. Adult Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in AV condition

Fixed Effects:	Estimate	Std. Error	z value	p
(Intercept)	1.78	0.14	12.18	< 0.001
SNR	0.32	0.01	22.62	< 0.001
Age group	-1.92	0.19	-9.65	< 0.001
SNR:Age group	0.0001551	0.01	0.008	0.994

Table 12. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in Pink noise in AV condition

Fixed Effects:	Estimate	Std. Error	z value	P
(Intercept)	1.96	0.17	11.45	< .001
SNR	0.16	0.01	14.08	< .001
Age group	-2.40	0.24	-10.02	< .001
SNR:Age group	0.15	0.01	9.09	< .001

Table 13. Results of the Linear Mixed Effects Logistic Regression on Intelligibility Data in two-talker babble in AV condition

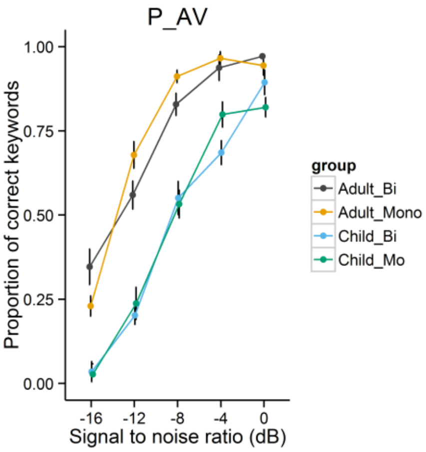

Figure 3. Performance in pink noise masker with visual cues.

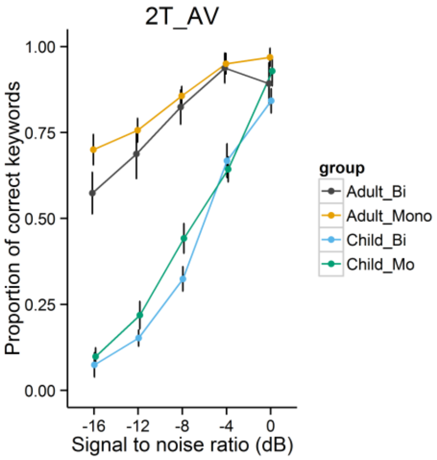

Figure 4. Performance in two-talker babble masker with visual cues.

Visual Enhancement The Wald test was used to test the overall effect and interaction. The analysis reveals a main effect of SNR (p<.001) and a main effect of age group (p=.04). However, since there are higher-order interactions with both of them, these two main effects are not interpreted. We found four different interactions, namely masker type × age group interaction (p=.01), SNR × age group (p<.001), SNR × masker type (p<.001), and SNR × masker type × language group (p=.01). It should be noted that there is no SNR × masker type × language group × age group interaction (p=.51; Table 14).

Fixed Effects:	Chi sq	Df	p
SNR	80.74	4	< .001
Masker type	2.12	1	.14510
Age group	4.11	1	.043
Language group	1.09	1	.29662
SNR:Masker type	36.50	4	< .001
SNR:Age group	60.8	4	< .001
Masker type:Age group	6.59	1	.010
SNR:Language group	2.53	4	.63917
Masker type:Language group	0.70	1	.40199
Age group:Language group	0.8	1	.77315
SNR:Masker type:Age group	4.94	4	.29329
SNR:Masker type:Language group	13.14	4	.011
SNR:Age group:Language group	1.78	4	.77680
Masker type:Age group:Language group	0.01	1	.92972
SNR:Masker type:Age group:Language group	3.27	4	.51302

Table 14. Wald test for main effect and interaction in Visual Enhancement

First we focus on teasing apart the masker type × age group interaction and SNR × age group interaction due to our primary interest on the developmental patterns of visual enhancement. Since there is no SNR × age group × language group interaction (p=.78), masker type × age group × language group interaction (p=.93), or 4-way interaction as mentioned above, there is no statistical evidence to support that the patterns as described below for masker type × age group interaction and SNR × age group interaction differ across monolinguals and bilinguals.

The masker type × age group interaction suggests that in pink noise the overall VE of adult's with all SNR collapsed is larger than that of child's (p<.002), yet the difference between both age groups in two-talker condition does not reach statistical significance (p=.83; Table 14). With regard to the SNR × age group interaction, further analysis of this interaction reveals that adult's VE is larger than that of the child's in more challenging listening conditions at -12 and -16 dB but not in other SNR levels (Table 15). There is no statistical evidence to support that this pattern differs across masker types since there is no SNR × age group × masker type interaction (p=.29).

	Estimate	Standard error	DF	t-value	Lower CI	Upper CI	p
2T:Age Group	0.0	0.0274	186.8	0.22	-0.0480	0.0599	.828
Pink:Age Group	0.1	0.0274	186.8	3.12	0.0313	0.1393	.002

Table 15. Breakdown of masker type × age group interaction

	Estimate	Standard error	DF	t-value	Lower CI	Upper CI	p
0 SNR:Adult:Child	0.0	0.0392	457.7	-1.20	-0.1239	0.0300	.231
-4 SNR:Adult:Child	-0.1	0.0392	457.7	-1.58	-0.1389	0.0150	.114
-8 SNR:Adult:Child	-0.1	0.0392	457.7	-1.62	-0.1405	0.0134	.105
-12 SNR:Adult:Child	-0.2	0.0392	457.7	4.37	0.0940	0.2479	< .001
-16 SNR:Adult:Child	0.2	0.0392	457.7	5.87	0.1528	0.3067	< .001

Table 16. Breakdown of SNR × age group interaction

Since there is a three-way SNR × masker type × language group interaction, the 2-way SNR × masker type interaction is not interpreted. Further breakdown of the 3-way interaction provides statistical evidence for the existence of different patterns of interactions between language groups with particular SNR levels in different maskers. In the pink noise condition, monolinguals displayed greater visual enhancement at -12dB (p=.006; Table 17; Figure 5). On the other hand, in two-talker babble, monolinguals displayed less visual enhancement at SNR -4 dB (p=.009; Table 18; Figure 6).

Fixed Effects:	Estimate	Std. Error	df	t value	p
(Intercept)	0.015	0.0398	287.5	.396	.692746
-4 SNR	0.0351	0.0539	235.9	0.65	.515223
-8 SNR	0.1862	0.0539	235.9	3.45	.000664
-12 SNR	0.2104	0.0539	235.9	3.89	.000126
-16 SNR	0.1392	0.0539	235.9	2.57	.010538
Language group	-0.0652	0.0533	287.5	-1.22	.221923
-4 SNR:Language group	0.0990	0.0723	235.9	1.369	.172275
-8 SNR:Language group	0.0969	0.0723	235.9	1.341	.181344
-12 SNR:Language group	0.1994	0.0723	235.9	2.758	.006275
-16 SNR:Language group	0.0457	0.0723	235.9	0.632	.527769

Table 17. Breakdown of SNR × language group in pink noise masker

Fixed Effects:	Estimate	Std. Error	df	t value	p
(Intercept)	-0.03951	0.04294	294.8	-0.920	.358312
-4 SNR	0.22701	0.06042	235.9	3.757	.000217
-8 SNR	0.13441	0.06042	235.9	2.225	.027050
-12 SNR	0.15988	0.06042	235.9	2.646	.008690
-16 SNR	0.17377	0.06042	235.9	2.876	.004396
Language Group	0.12590	0.05752	294.8	2.189	.029381
-4 SNR:Language group	-0.21230	0.08093	235.9	-2.623	.009276
-8 SNR:Language group	-0.11052	0.08093	235.9	-1.366	.173363
-12 SNR:Language group	-0.09921	0.08093	235.9	-1.226	.221438
-16 SNR:Language group	-0.03038	0.08093	235.9	-0.375	.707677

Table 18. Breakdown of SNR × language group in two-talker masker

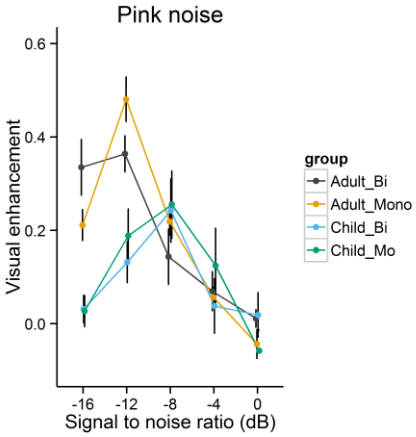

Figure 5. Visual enhancement in pink noise.

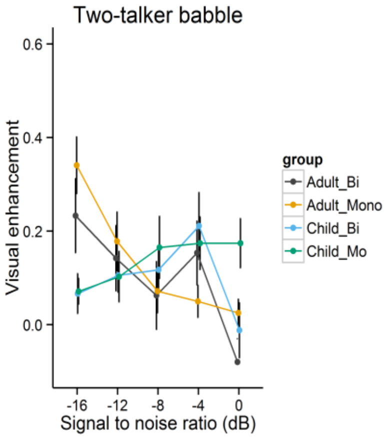

Figure 6. Visual enhancement in two-talker babble.

DISCUSSION

This project investigated the extent to which the age and language background of the listener modulated maximal intelligibility benefits from audiovisual integration. To achieve this goal, the impact of audiovisual processing on intelligibility was examined across a range of SNRs (0 to -16 dB) in an energetic masker, pink noise condition, and a two-talker babble condition, which is primarily a type of informational masker, however small amounts of energetic masking are still present (Brungart et al., 2001). The described conditions were utilized for the presentation of English sentences produced by a native male, American English speaker to four groups of listeners: monolingual and bilingual native English children, and monolingual and bilingual native English adults.

Based upon the gain in speech perception-in-noise performance in the AO condition compared to significant differences found in the AV condition, it can be concluded that all groups rely on audiovisual modalities to enhance intelligibility in adverse listening conditions. These results are consistent with previous findings that also demonstrate an increase in intelligibility when speech stimuli are presented in an AV condition (Helfer & Freyman, 2005; Ross et al., 2011).

Although audiovisual speech perception resulted in benefited speech intelligibility, the same increase in intelligibility was not observed for all groups. Both monolingual and bilingual children exhibited an increased visual enhancement at easier SNRs, while adult groups demonstrated increased visual enhancement at more intermediate SNRs (according to Ross et al., 2007) in both masking conditions. These results suggest that adults have more advanced audiovisual integration and are therefore able to benefit more from visual articulatory cues in more severe adverse listening conditions. One explanation for observed differences in adult and child performance is due to the child’s limited language experience (Elliot, 1979). However, this explanation can be dismissed as all target words in this experiment were screened to ensure that they were developmentally appropriate for children in our age range. Ross et al. (2007) found a significant increase in AV performance from the young child group (age range=5-7) when compared to a slightly older group (age range=8-9); however, they found very little difference in AV gain from the 8-9 year group compared to the 10-11 year group. The authors additionally found that a significant increase in AV gain in the 12-14 group, which was similar to adult performance. Based upon these results, in a future analysis we aim to observe the difference between the current study’s child groups 6-7 (n= 19) and 8-10 (n=11), to investigate a more fine-grained developmental influence.

In regard to masking conditions, a clear difference was noted as children showed higher performance in pink noise than in two-talker babble, while adults showed higher accuracy in two-talker babble when compared to their performance in the pink noise condition in both AO and AV conditions. This may be due to the fact that the children have not fully developed cognitive compensatory factors such as working memory and attention (Wightman & Allen, 1992). The better performance in adults in the two-talker babble condition replicates previous findings, which indicate that two-talker babble results in a limited amount of energetic masking, but because speech is redundant, listeners can in turn perceive glimpses to recognize target speech (Cooke, 2006). This serves as another piece of supporting evidence for the child’s emergent cognitive compensatory factors. That is, the child may not be able to take advantage of adult-like *glimpsing* in order to attend to and perceive salient phonemic information because that skill has not fully developed.

In regard to language factors, bilingual children perform more similarly to their monolingual counterparts than bilingual adults. Based on the results of this study, monolingual and bilingual children did not differ significantly on their performance in the SPIN task. This finding is in contrast to the past studies investigating speech perception-in-noise performance in bilingual children and their monolingual counterparts. This could be due to the fact that the bilingual child group in the present study all had a simultaneous onset of their second language. Moreover, each participant had a high proficiency and daily usage of both of their languages. Recall that the majority of past research conducted studies on non-native adult participants who acquired their second language before age 6 (Rogers et al., 2006; Tabri, Chacra, & Pring, 2011). The similar performance found in the child monolingual and simultaneous, highly proficient bilingual child participants may indicate that there is a sensitive period in development when bilinguals can perform as well as monolinguals on speech perception-in-noise tasks. Monolingual adults exhibited a steeper peak for visual enhancement at -12dB SNR, replicating Ross et al.'s findings of a window of maximal multisensory integration beyond the predictions of the principle of inverse effectiveness. These results may indicate that the divergent use of visual cues in speech perception between bilingual and monolingual speakers occurs later in development. Therefore, the results for only the monolingual adults support the intermediate zone hypothesis, which predicts maximal intelligibility gain for intermediate SNRs.

CONCLUSION AND FUTURE IMPLICATIONS

Visual cues enhance speech perception in both energetic and informational masking conditions across all groups. However, the amount of benefit from audiovisual integration differed across the two types of maskers, in both child and adult participants. In energetic masking, for adult monolingual participants the visual gain in speech intelligibility is maximal at intermediate SNR (-12 dB). This was not found in bilingual adult participants. In contrast, in informational masking, the visual gain in speech intelligibility increased as SNRs became more difficult and was maximal at the most difficult SNR (-16 dB). Therefore, speech perception in informational masking is consistent with the principle of PoIE (Sumby & Pollack, 1954; Erber, 1969; Meredith & Stein, 1986), while speech perception in energetic masking for monolingual adults follows the window of maximal audiovisual integration theory (Ross et al., 2007; Ross et al., 2011). However, this pattern was not found in bilingual adults, nor in the two child groups. In contrast, children showed higher performance in pink noise than in two-talker babble in both AO and AV conditions.

Due to the heterogeneity of the student population, it is a challenge to fully understand the nature of individual differences found in the developmental modulation of auditory and visual processing in speech perception. However, the findings here present statistical evidence for the ongoing development of the fusion of audiovisual modalities and the benefit of visual cues during speech perception in adverse listening conditions. With the current knowledge that is available regarding the salience of visual cues to enhance speech perception-in-noise, future studies should continue exploring multisensory processing in children and adults, implementing supplementary non-linguistic attention and executive function tasks, as well as neural tasks.

Appendix

1. The hot sun warmed the ground.
2. The gray mouse ate the cheese.
3. The strong father carried my brother.
4. The large monkey chased the child.
5. The mean bear ate the fruit.
6. The loud noise upset the baby.
7. The friendly neighbor helped the grandmother.
8. The black bear scared the visitors.
9. The hungry children ate the snacks.
10. The strong sister won the game.
11. The rude joke upset my parents.
12. The dark house scared the baby.
13. The talented musician knew the songs.
14. The gray horse ate the grass.
15. The sick student read the book.
16. The hungry girl made the sandwich.
17. The tiny flies bothered the girl.
18. The new student liked the professor.
19. The hot coffee hurt the boy.
20. The small animal scared the baby.
21. The teacher chose the horrible book.
22. The children enjoyed the holiday parade.
23. The girl loved the sweet coffee.
24. The grandmother baked a sweet cake.
25. The woman met the rich actor.
26. The doctor owned the yellow car.
27. The teacher wrote a difficult question.
28. The store sold the dirty clothes.
29. The ball broke the glass window.
30. The grandfather loved the red wine.
31. The brother met the talented artist.
32. The chef baked the sweet corn.
33. The father hugged his sad daughter.
34. The chef cooked the delicious food.
35. The bird found the juicy worm.
36. The grandfather drank the dark coffee.
37. The neighbor liked the loud song.
38. The cat chased the gray mouse.
39. The mother baked the delicious cookies.
40. The team played a difficult game.
41. The kind girl helped the strangers.
42. The talented author received the prize.
43. The black cat climbed the tree.
44. The thoughtful boyfriend bought the flowers.
45. The hungry dog ate the food.
46. The friendly cat loved the boy.
47. The old man cooked the carrots.
48. The happy dog found the toy.
49. The youngest sisters watched the parade.
50. The sweet dog found the toy.
51. The pretty girl won the prize.
52. The lonely artist called her friend.
53. The youngest child hated the fruit.
54. The cheap food attracted the customers.
55. The rich boyfriend owned the houses.
56. The new kitten climbed the tree.
57. The angry bear scared the couple.
58. The thirsty cat drank the milk.
59. The three sisters shared the clothes.
60. The tiny rabbit chewed the grass.
61. The wind destroyed the tiny house.
62. The restaurant sold the red wine.
63. The musician played a beautiful song.
64. The boy carried the heavy chair.
65. The chef chose the delicious cheese.
66. The man ate the large meal.
67. The parents told the horrible story.
68. The man shared the difficult story.
69. The chef made the fresh noodles.
70. The teacher read an interesting novel.
71. The restaurant served a delicious soup.
72. The woman heard a beautiful song.
73. The grandmother loved the rich cake.
74. The nurse cleaned the dirty clothes.
75. The family watched the talented performer.
76. The author told an interesting story.
77. The painter owned the soft brushes.
78. The store sold the delicious food.
79. The travelers visited the new museum.
80. The bird bothered the old dog.

References

Alcántara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107-1114.

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, *2*(1), 42-46.

American Speech-Language Hearing Association. (1978). Guidelines for manual pure-tone threshold audiometry. *ASHA*, *20*, 297-301.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2013). Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *Journal of Speech, Language, and Hearing Research*, *56*(1), 31-43.

Arbogast, T. L., Mason, C. R., & Kidd Jr, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, *112*(5), 2086-2098.

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology*,*41*(3), 541.

Baker, C. (1993). *Foundations of bilingual education and bilingualism*. Clevedon, England: Multilingual Matters.

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults.*Journal of experimental child psychology*, *105*(1), 38-50.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *125*(5), 3358-3372.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3-11.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging*, *19*(2), 290.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental science*, *7*(3), 325-339.

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognitive Development 11*(4), 485-508.

Bishop, D. V., & McArthur, G. M. (2005). Individual differences in auditory processing in specific language impairment: A follow-up study using event-related potentials and behavioural thresholds. *Cortex*, *41*(3), 327-341.

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, *118*(2), 245-257.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05)[Computer program].

Bovo, R., & Callegari, E. (2009). Effects of classroom noise on the speech perception of bilingual children learning in their second language: Preliminary results. *Audiological Medicine*, *7*(4), 226-232.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339-2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272-284.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex*, *21*(5), 1042-1055.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527-2538.

Calandruccio, L., & Smiljanic, R. (2012). New sentence recognition materials developed using a basic non-native english lexicon. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1342-1355.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562-1573.

Cooke, M., Lecumberri, M. G., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America,123*(1), 414-427.

Cummins, J. (1976). The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses. Working Papers on Bilingualism, No. 9.

Crandell, C. C., & Smaldino, J. J. (1996). Speech perception in noise by children for whom English is a second language. *American Journal of Audiology*, *5*(3), 47.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, *34*(3), 565-571.

Dawes, P., & Bishop, D. V. (2008). Maturation of visual and auditory temporal processing in school-aged children. *Journal of Speech, Language, and Hearing Research*, *51*(4), 1002-1015.

Davis, J., & Bauman, K. “School Enrollment in the United States: 2011,” *Population Characteristics*, P20-571, U.S. Census Bureau, September 2013, <http://www.census.gov/prod/2013pubs/p20-571.pdf>

Diaz, R. M. (1983). Thought and two languages: The impact of bilingualism on cognitive development. *Review of research in education*, 23-54.

Dijkstra, T., & Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. *Localist connectionist approaches to human cognition*, 189-225.

Dijkstra, A. F. J., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision.

Eggermont, J. J. (1985). Physiology of the developing auditory system. In *Auditory development in infancy* (pp. 21-45). Springer US.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*, *107*(5), 2704-2710.

Elliot, J. J. (1988). Physiology of the developing auditory system. In S. E. Trehub & B. A. Schneider (Eds.), Auditory development in infancy (pp. 21-45). New York: Plenum.

Elliot, 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,* 651-653.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*(2), 423.

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, *40*(4), 481-492.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *The Journal of the Acoustical Society of America*, *108*(6), 3023-3029.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578-3588.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism Language and Cognition*, *6*(1), 1-15.

Hall, J. W., Buss, E., Grose, J. H., & Dev, M. B. (2004). Developmental effects in masking-level difference. *Journal of Speech, Language, and Hearing Research, 47*, 13-20.

Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, *33*(1), 149-155.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, *117*(2), 842-849.

Hétu, R., Truchon-Gagnon, C., & Bilodeau, S. A. (1990). Problems of noise in school settings: A review of literature and the results of an exploratory study. *Journal of Speech-Language Pathology and Audiology*, *14*(3), 31-39

Hodgson, M. (2002). Rating, ranking, and understanding acoustical quality in university classrooms. *The Journal of the Acoustical Society of America*,*112*(2), 568-575.

Jansen, S., Chaparro, A., Downs, D., Palmer, E., & Keebler, J. (2013, September). Visual and Cognitive Predictors of Visual Enhancement in Noisy Listening Conditions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1199-1203). SAGE Publications.

Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children’s sensitivity to visual speech: A new multimodal picture–word task. *Journal of experimental child psychology*, *102*(1), 40-59.

Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language & Hearing Research*, *43*(1), 144-157.

Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39,* 159-207.

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Brief Intelligence Test, Second Edition. Bloomington, MN: Pearson, Inc.

Kessler, C., & Quinn, M. E. (1987). Language minority children's linguistic and cognitive creativity. *Journal of Multilingual & Multicultural Development*, *8*(1-2), 173-186.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*(1), 110-124.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599-605.

Kroll, J. F., Gullifer, J. W., & Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annual Review of Applied Linguistics*, *33*, 102-127.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138-1141

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. *Advances in Psychology*, *134*, 59-85.

Li, P., Zhang, F., Tsai, E., Puls, B. (2013). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, DOI: 10.1017/S1366728913000606.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One*, *4*(3), e4638.

Marian, V. (2009). Audio-visual integration during bilingual language processing. *The bilingual mental lexicon: Interdisciplinary approaches*, 52-78.

Mattys, S. L., Carroll, L. M., Li, C. K., & Chan, S. L. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11), 887-899.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953-978.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686.

McLaughlin, B. (Ed.). (2013). *Second Language Acquisition in Childhood: Volume 2: School-age Children*. Psychology Press.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*(3), 640-662.

McLeod, S. (Ed.). (2007). *The international guide to speech acquisition*. Clifton Park, NY: Thomson Delmar Learning.

Moore, J. K. (2002). Maturation of human auditory cortex: Implications for speech perception. *Ann Otol Rhinol Laryngol.*

Navarra, J., Yeung, H. H., Werker, J. F., & Soto-Faraco, S. (2012). Multisensory interactions in speech perception. In B. E. Stein (Ed.), *The New Handbook of Multisensory Processing* (pp. 435-452). Cambridge, MA: MIT Press.

Nelson, P., Kohnert, K., Sabur, S., & Shaw, D. (2005). Classroom noise and children learning through a second language: double jeopardy?. *Language, Speech & Hearing Services in Schools*, *36*(3).

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, *87*(6), 2705-2715.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic approaches to bilingualism*, *1*(3), 213-237.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191-196.

Pavlenko, A. (Ed.). (2009). *The bilingual mental lexicon: Interdisciplinary approaches* (Vol. 70). Multilingual Matters.

Picard, M., & Bradley, J. S. (2001). Revisiting Speech Interference in Classrooms: Revisando la interferencia en el habla dentro del salón de clases. *International Journal of Audiology*, *40*(5), 221-244.

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. Clinical Neurophysiology, 111(2), 220-236.

Riley, K. G., & McGregor, K. K. (2012). Noise hampers children's expressive word learning. *Language, speech, and hearing services in schools*, *43*(3), 325-337.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(03), 465-485.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147-1153.

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. *European Journal of Neuroscience*, *33*(12), 2329-2337.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. *Handbook of child psychology*.

Schneider, W., Eschman, A., Zuccolotto, A., & Guide, E. P. U. S. (2002). Psychology Software Tools Inc. *Pittsburgh, USA*.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), B69-B78.

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11*(2), 306-320.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., ... & Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience*, *28*(14), 3586-3594.

Sheng, L., Lu, Y., Kan, P. (2011). Lexical development in Mandarin–English bilingual children. *Bilingualism: Language and Cognition*, *14,* 579–587

Shimizu, T., Makishima, K., Yoshida, M., & Yamagishi, H. (2002). Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx*, *29*(2), 121-125.

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *The Journal of Neuroscience*, *24*(38), 8223-8231.

Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and hearing*, *26*(1), 78-88.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212-215.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 335(1273), 71-78.

Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, *225*(1), 117-127.

Tabri, D., Chacra, K. M. S. A., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders*, *46*(4), 411-422.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language learning and development*, *3*(1), 73-100.

U.S. Census Bureau, 2009 American Community Survey, B16001, "Language Spoken at Home by Ability to Speak English for the Population 5 Years and Over," <http://factfinder.census.gov/>, accessed January 2011.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519-526.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first-and second-language multi-talker babble. *Speech communication*, *52*(11), 943-953.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510-540.

Van Heuven, W. J., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews, 64*(1), 104-122.

Von Hapsburg, D., Champlin, C. A., & Shetty, S. R. (2004). Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *Journal of the American Academy of Audiology, 15*(1), 88-98.

Von Hapsburg, D., & Bahng, J. (2006). Acceptance of background noise levels in bilingual (Korean-English) listeners. *Journal of the American Academy of Audiology*, *17*(9).

Werner, L. A. (2007). What do children hear: How auditory maturation affects speech perception. *The ASHA Leader, 12*(6-7), 32-33.

Wightman, F., & Allen, P. (1992). Individual differences in auditory capability among preschool children. *Developmental Psychoacoustics,* 113-133

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languagesa). *The Journal of the Acoustical Society of America*, *123*(6), 4524-4538.

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: clinical implications. *International Journal of Language & Communication Disorders, 44*(3), 253-270.

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants’ audiovisual speech perception. *Psychological science*, *24*(5), 603-612.