DEVELOPMENTAL CHANGES IN THE TEMPORAL WINDOW OF AUDITORY AND VISUAL INTEGRATION

By

Andrea R. Hillock

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing & Speech Sciences

August, 2010

Nashville, Tennessee

Approved:

Professor Mark Wallace

Professor Linda Hood

Professor Wesley Grantham

Professor Bruce McCandliss

To my Mom and Dad, whose love, faith and support is beyond compare,

and

to my wonderful husband, James, who makes me smile everyday

ACKNOWLEDGEMENTS

This work would not have been possible without the generous financial support from the Vanderbilt Kennedy Center and the American Speech Language and Hearing Foundation.

Special thanks are also due my partners in crime, the students and postdocs (current & alumni) from the Wallace lab (Al, Leslie, Vera, Juli, Diana, Matt, Dips, Caitlin, Aaron, Brian and David), Hood lab (Chris and Erin) and my cohort in the Hearing & Speech Sciences program that shared the ups and downs of this journey with me. Al Powers has been a confidant since I joined the Multisensory Processes Lab and I have thoroughly enjoyed our continued collaboration on all things PSS. I'd also like to especially thank Leslie, my go-to Stats guru, and Vera for her unending willingness to dialogue about multisensory development. I would be remiss if I did not also mention the students and staff that assisted with subject scheduling, data collection and analysis (Michelle Young, Nicky Hackett, Olivia Broaddus, Matthew Hevey, Haley Eberly), as well as Jurnell Cochkren and Zachary Barnett for their technical assistance.

Lastly, I want to express my gratitude to my dissertation committee. To Dr. McCandliss, whose critical feedback during the prospectus and writing process substantially elevated this project, Dr. Grantham, who taught me about staircases, SDT, MATLAB and helped troubleshoot countless bugs over the last 4 years, and to Dr. Hood, for providing me with ANSD opportunities, guidance and advice on all of my projects and career steps. I also want to sincerely thank my mentor and committee chair, Dr. Mark Wallace, who has always made time to field my questions and challenge my thinking. I greatly appreciate both his commitment to my growth as a future scientist and his instructive comments on all of my work. I feel privileged to have been his mentee and I look forward to our continued friendship and collaboration in the years to come!

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

GENERAL INTRODUCTION

Prologue

"Humans live in a world of objects and events that can be seen, heard, and felt" (Spelke, 1979, p. 626). Combining information from multiple senses can profoundly affect perception and behavior even in early postnatal life. Such multisensory effects are strongly influenced by the temporal relationship between stimuli, and while this has been well characterized in adults, very little is known about multisensory temporal processing in the period between infancy and adulthood. Thus, the overarching goal of this thesis is to examine maturation of the temporal structure of audiovisual binding for speech and nonspeech signals across an extended period of development. The sections to follow will review the literature germane to the development of multisensory integration and audiovisual temporal processing.

First, the byproducts of multisensory interactions will be described to introduce various methods of indexing integration and to reinforce the powerful influence of these effects on our behavior and perception. Next, the influence of spatial and temporal congruity and stimulus strength on multisensory interactions will be reviewed, with special attention given to the temporal aspects of multisensory processing. Finally, the latter portion of this introduction will focus on age-related changes in unisensory and multisensory processing, and the potential role of experience in mediating maturation of multisensory (temporal) abilities.

Establishing the trajectory for the development of multisensory temporal processing is critically important given the influence of audiovisual integration for processes as diverse as reading, learning and the perception of speech and nonspeech environmental stimuli. The realization of mature processing has global implications for the child, enabling behavioral enhancements that span academic, communication/ social and motor realms.

Behavioral Consequences of Multisensory Integration

The environment is replete with sensory information and the brain has evolved to manage the ongoing barrage of stimulation. It effectively integrates independent sensory signals to produce comprehensive representations of the objects and events in our world. This process not only improves our ability to detect and resolve stimuli, but has been shown to profoundly alter perception.

Multisensory illusions have provided particularly compelling examples of the ability of one stimulus to modulate perception of another. A classic example, the Ventriloquist effect, refers to the spatial mislocalization of an acoustic stimulus (Howard & Templeton, 1966). To induce the illusion, an actor restricts his/her lip movement and temporally synchronizes his/her utterances to the opening and closing of a puppet's mouth, thus producing the perception that the puppet is speaking. An equally convincing speech-based multisensory illusion is the McGurk effect, where presentation of incongruent auditory (e.g., /ba/) and visual (e.g., /ga/) tokens frequently results in perception of a fused intermediary token (e.g., /da/, /tha/) (McGurk & MacDonald, 1976). Basic stimuli have also been used to produce robust crossmodal illusions such as the flash-beep effect in which pairing a single flash with multiple beeps can result in the perception of multiple flashes (Shams, Kamitani & Shimojo, 2002).

While the use of conflicting cues can provide a window into the interplay between sensory systems, the facilitative effects of congruent bimodal stimulation hold extreme ethological importance and can also be used to uncover multisensory interactions. Studies have reported marked facilitations of response times (Frens, Van Opstal & Van der Willigen, 1995; Hecht, Reiner & Karni, 2008; Hershenson, 1962; Hughes, Reuter-Lorenz, Nozawa & Fendrich, 1994), and improvements in stimulus detectability during multisensory compared to unisensory stimulation (Frassinetti, Bolognini & Ladavas, 2002; Lovelace, Stein & Wallace, 2003; Stein, Huneycutt & Meredith, 1988). For example, cats trained to orient to a dim light show improved localization abilities (increased accuracy) in conditions when the light was paired with an irrelevant sound (Stein et al., 1988). Multisensory benefits have also been reported on measures assaying discrimination of auditory speech. Seminal work by Sumby and Pollack showed that recognition of speech in noise is enhanced in audiovisual relative to unisensory conditions (Sumby & Pollack, 1954). Subsequently, Macleod and Summerfield quantified this effect, reporting differences in auditory only and audiovisual speech reception thresholds (SRTs) that were consistent with an effective 11 dB increase in signal-to-noise ratio (SNR) under multisensory conditions (Macleod & Summerfield, 1987). Findings reiterate the competitive advantage conferred by multisensory processing, and illustrate that stimuli from one sensory modality can alter, strengthen, or accelerate the response to an associated stimulus from another modality.

Principles of Multisensory Integration

Given that the brain must simultaneously process many sensory events, how does it discern what stimuli are related and should be integrated? Several principles have emerged from unisensory physiology studies in animals that have since been applied to humans which describe techniques used by the brain to make such determinations. These principles are critically important for establishing the likelihood of interactions and subsequent behavioral effects.

The Spatial Principle

The pioneering work on the influence of interstimulus relationships on multisensory processing comes from single unit recordings in the superior colliculus (SC) of cats, a structure in the midbrain that receives input visual, auditory and somatosensory input (Stein & Arigbeded, 1972). The spatial receptive fields of multisensory cells in the intermediate and deep layers of the SC largely overlap across sensory modalities (Meredith & Stein, 1986a). Multisensory enhancements (i.e., neuronal discharge to multisensory stimuli exceeding the best unisensory response) are observed when stimuli are presented within the same receptive field (Meredith & Stein, 1996), whereas presentation of stimuli in non-overlapping receptive fields often leads to response depressions (i.e., decreases in firing relative to the most effective unisensory response) (Meredith & Stein, 1996). Hence, because of the spatial overlap of neuronal receptive fields, multisensory stimuli that originate from or near the same location give rise to large response enhancements.

Spatial congruence can also substantially affect multisensory enhancements in humans. Facilitation of saccadic reaction times is reduced as the spatial distance between paired stimuli is increased (Amlot, Walker, Driver & Spence, 2003; Colonius & Arndt, 2001; Corneil & Munoz, 1996; Harrington & Peck, 1998), and performance on audiovisual temporal tasks is improved when stimuli are presented from different positions (Keetels & Vroomen, 2005; Zampini, Guest, Shore & Spence, 2005b; Zampini, Shore & Spence, 2003), suggesting that the brain is less likely to bind more spatially segregated signals. Such a strategy is quite intuitive and effectual in preventing integration of unrelated sensory stimuli, given that in real world scenarios stimuli arising from distinct locations in the environment are unlikely to be associated with one another.

The Principle of Inverse Effectiveness

Multisensory interactions are not only shaped by spatial factors, but are strongly modulated by the effectiveness of the individual sensory cues. Numerous studies have illustrated that the magnitude of multisensory interactions is inversely proportional to the effectiveness of the unisensory stimuli (Meredith & Stein, 1986b). In other words, maximal enhancement is produced by the pairing of two weakly effective unisensory stimuli, whereas highly effectively unisensory stimuli are unlikely to produce robust facilitation (Meredith & Stein, 1986b). Findings have important implications on behavior, suggesting that multisensory integration is most robust in situations where gains are most needed - those in which unisensory stimuli are degraded or difficult to decipher. Consistent with this, decreases in visual benefit on audiovisual speech recognition measures are reported with increasing SNR (Erber, 1969; Sumby & Pollack, 1954). Effects are supported with neuroimaging work showing decreases in the magnitude of facilitation in multisensory regions to audiovisual speech presented in higher levels of noise relative to low noise and quiet (Callan et al., 2003; Stevenson & James, 2009), but see Ross et al. (2007) for evidence suggesting that maximal enhancements for speech are attained at intermediary SNRs.

The Temporal Principle and the Multisensory Temporal Binding Window

Like spatial factors and stimulus effectiveness, temporal congruence is also a critical determinant of multisensory integration. Physiology research in cats indicates that multisensory enhancements are generally observed for SOAs of approximately 200 ms or less, but that enhancement is maximal when the peak discharge periods to unisensory stimuli overlap (Meredith, Nemitz & Stein, 1987).

These findings have been replicated in human behavior and neuroimaging work, where the temporal principle has been expanded to include the concept of a “temporal binding window”. This window characterizes the span of time over which interactions are likely to be generated. Multisensory stimuli presented in close temporal proximity can produce behavioral changes (e.g., improved speech recognition, speeded reaction times, improved saccadic accuracy); however, sufficient temporal misalignment minimizes effects. For example, gains in speech perception in noise from the addition of visual cues are typically preserved for SOAs up to about 100 ms after which time further delay reduces multisensory benefit (McGrath & Summerfield, 1985; Pandey, Kunov & Abel, 1986). Complementary findings are reported in studies characterizing the perseverance of illusions such as the McGurk effect, where fusions indicative of interactions reportedly persist over asynchronies of 150 to 200 ms or more, but are decreased at larger stimulus onset asynchronies (SOAs) (Jones & Jarick, 2006; Massaro, Cohen & Smeele, 1996; Munhall, Gribble, Sacco & Ward, 1996; van Wassenhove, Grant & Poeppel, 2007). Similar reductions in illusion perception with increasing intersensory temporal disparity have been identified with the flash-beep (Shams et al., 2002) and stream-bounce illusions (Sekuler, Sekuler & Lau, 1997).

Differences in behavioral report or illusion perception for synchronous relative to asynchronous presentations have been linked to differential activity in a number of subcortical and cortical regions including the superior colliculus (Bushara, Grafman & Hallett, 2001; Calvert, Hansen, Iversen & Brammer, 2001), superior temporal sulcus (Calvert et al., 2001; Dhamala, Assisi, Jirsa, Steinberg & Kelso, 2007; Noesselt et al., 2007; Powers, Hevey & Wallace, Unpublished results), posterior parietal cortex (Bushara et al., 2001; Dhamala et al., 2007; Powers et al., Unpublished results), insula (Bushara et al., 2001; Calvert et al., 2001; Powers et al., Unpublished results), frontal (Calvert et al., 2001) and prefrontal regions (Bushara et al., 2001; Dhamala et al., 2007) and primary sensory cortices (Dhamala et al., 2007; Noesselt et al., 2007; Powers et al., Unpublished results). Responses are generally greater for synchronous presentations, although increased activity in the left occipito-temporal junction and superior colliculus has been reported for asynchronous audiovisual pairings (Dhamala et al., 2007; Jones & Callan, 2003).

Interestingly, the influence of interstimulus temporal synchrony on multisensory integration appears to differ across stimuli (e.g., speech, nonspeech) and according to the order in which sensory signals are presented (e.g., auditory or visual first). Using a method of constant stimulation and incremental adjustment, Dixon and Spitz (1980) derived threshold values in adults indicative of the delay at which they were first able to detect asynchrony in auditory and visual tracks of movies displaying a hammer hitting a board and a person reading. Mean asynchrony detection thresholds were higher for speech than nonspeech presentations, a finding supported by comparisons of window size across recent studies (Powers, Hillock & Wallace, 2009; van Wassenhove et al., 2007). In addition, a stimulus order effect was observed. Thresholds were elevated for visual leading relative to visual lagging temporal delays, an effect that has also since been corroborated (Fujisaki, Shimojo, Kashino & Nishida, 2004; Jones & Jarick, 2006; Massaro et al., 1996; Munhall et al., 1996; Powers et al., 2009; van Wassenhove et al., 2007; Vroomen, Keetels, de Gelder & Bertelson, 2004; Zampini, Guest, Shore & Spence, 2005a). This suggests that participants detect asynchrony less readily when visual stimuli appear first.

Thus, it has been shown that the size of the temporal binding window can vary, but is the window derived to a given stimulus combination fixed? Research on multisensory temporal processing has begun to examine the plasticity of the temporal binding window in studies assaying the effects of recalibration and training in adults. Brief and repeated exposure to asynchronous audiovisual stimuli reportedly shifts the point of subjective simultaneity on temporal order (Vroomen, 2004) and simultaneity judgment (Fujisaki et al., 2004) tasks in the direction of the exposed lag. Moreover, recent work from our lab has indicated that training (with feedback) can produce lasting decreases in the width of the temporal binding window (Powers et al., 2009), and that such changes are correlated with alterations in activity in a network of multisensory and unisensory areas (Powers et al., Unpublished results). Findings are extraordinarily exciting and suggest the potential for multisensory based training to be used as a tool for remediating processing alterations in clinical populations where enlarged windows have been previously reported (e.g., dyslexia, autism) (Foss-Feig et al., 2010; Hairston, Burdette, Flowers, Wood & Wallace, 2005).

Maturation of Unisensory Perception

The mechanism(s) driving window plasticity in adults may overlap with those mediating developmental changes in multisensory temporal processing; however, gains in unisensory functioning may also contribute to age-related changes in the multisensory binding window. Hence, to gain a fuller understanding of multisensory development, it is essential to explore maturational changes in the sensory systems contributing to the multisensory percept. Accordingly, what follows is a cursory review of some key benchmarks in visual and auditory development that may contribute to multisensory task performance in the studies contained in this thesis.

Visual System Development

While humans are fairly precocial, vision is significantly underdeveloped at birth. Visual acuity improves dramatically in the first six months of life, with more gradual changes continuing thereafter (Sireteanu, Fronius & Constantinescu, 1994; Teller, McDonald, Preston, Sebris & Dobson, 1986; Teller & Movshon, 1986). Measures of grating acuity (i.e., the ability to resolve a pattern of alternating dark and light stripes) in newborns suggest that their ability to resolve an object at 20 feet is comparable to that of adults at a viewing distance of 400 feet, but visual acuity improves quite rapidly reaching adult levels in some estimates by 3 years (Teller et al., 1986) and in others by 5-6 years (Birch, Gwiazda, Bauer, Naegele & Held, 1983). Substantial gains in dynamic vision perception (i.e., identification of object attributes via coherent motion of dots) and pattern perception (i.e., the capacity to resolve fine grained detail) are also observed in early postnatal development, although the time course to mature functioning is more protracted. Adult-like sensitivity emerges around 7 years for dynamic visual acuity (Giaschi & Regan, 1997; Parrish, Giaschi, Boden & Dougherty, 2005) and between 7 and 9 years of age for discrimination of static patterns (Adams & Courage, 2002; Ellemberg, Lewis, Hong Liu & Maurer, 1999; Gwiazda, Bauer, Thorn & Held, 1997).

Another important aspect of visual development that is particularly relevant to the current thesis is visual temporal resolution. Flicker-fusion studies have been useful in elaborating changes in visual temporal acuity with age by measuring one’s ability to detect interruptions in consecutive visual stimulation. Gaps that exceed temporal resolution thresholds can induce the perception of flicker, but this is eliminated when gap duration is shorter than human sensitivity thresholds. For example, movies are comprised of a series of consecutive images (frames) that are presented in close temporal succession such that observers do not perceive breaks in the film. Like other forms of visual perception, temporal acuity is reduced at birth and improves with age (Hartmann & Banks, 1992; Rasengane, Allen & Manny, 1997; Regal, 1981). Comparison of visual flicker-fusion thresholds in children and adults suggests that thresholds are mature for higher temporal frequencies before 4 years of age and for lower temporal frequencies around 7 years of age (Ellemberg et al., 1999; Hartmann, 1934). Thus, while substantial gains in visual acuity, contrast sensitivity and temporal resolution are reported in early life, maturational changes continue into early to middle childhood. Such changes are presumed to reflect foveal maturation, changes in system connectivity, and increasing axonal myelination, some of which may also contribute to delayed maturation of auditory and audiovisual functioning.

Auditory System Development

Unlike vision, the auditory system is functional even before birth. Responses to sound are consistently observed by the start of the third trimester when the pathway conducting sound from the ear to the brain is fully intact (Moore & Linthicum, 2007). Despite prenatal experience with sound and the presence of robust responses to audible stimuli, a number of auditory abilities continue to change through middle childhood and into adolescence (e.g., auditory sensitivity, frequency resolution and temporal resolution). Auditory thresholds, representing the minimum signal level required for stimulus detection, are elevated in infants (Nozza & Wilson, 1984; Olsho, Koch, Carter, Halpin & Spetner, 1988; Sinnott, Pisoni & Aslin, 1983), and reportedly continue to improve between 3 and 12 years of age (Jerger & Jerger, 1982; Maxon & Hochberg, 1982; Trehub, Schneider, Morrongiello & Thorpe, 1988). Interestingly, babies and children not only need louder presentations to be able to detect signals, but exhibit increased sensitivity to noise; they require higher SNRs to achieve comparable listening performance to adults (Buss, Hall, Grose & Dev, 1999; Hall, Grose, Buss & Dev, 2002; Jerger & Jerger, 1982). Differences in the ability to resolve pitch (frequency alterations) have also been identified in infants and children relative to adults (Hall & Grose, 1991; Olsho, 1984; Olsho, Koch & Halpin, 1987; Spetner & Olsho, 1990; Veloso, Hall & Grose, 1990). Whereas frequency selectivity is poorer in children under 4 years of age, performance appears adult-like in 6 year olds (Hall & Grose, 1991; Veloso et al., 1990).

Compared to frequency resolution, the timeline for maturation of auditory temporal acuity is slightly more extended. Gap detection thresholds (i.e., the smallest interval needed for participants to resolve a break in consecutive auditory presentations) are reportedly poorer in infants and young children relative to adults, although findings regarding the time frame for maturation of auditory temporal acuity are somewhat discrepant (Irwin, Ball, Kay, Stillman & Rosser, 1985; Trehub, Schneider & Henderson, 1995; Wightman, Allen, Dolan, Kistler & Jamieson, 1989). In a task comparing gap detection using broadband noise in 6 to 12 year olds and adults, Irwin et al. (1985) observed improvements from 6 to 10 years of age and the emergence of adult-like performance by age 11. The timeline for development depends on the spectral composition of the noise, with mature temporal resolution appearing around 6 years for high frequencies and 10 years for low frequencies (Grose, Hall & Gibbs, 1993; Irwin et al., 1985), but see Wightman et al. (1989) for opposite effects. Thus, findings suggest that whereas frequency discrimination abilities are mature by about the 1st grade, changes in auditory sensitivity and temporal resolution continue into middle childhood. As with the maturation of the visual system, these prolonged changes are likely the result of ongoing alterations in system connectivity and myelination.

Our survey of the unisensory literature suggests that a number of skills continue to mature into early to middle childhood, and that visual and auditory functioning may potentially influence the trajectory for audiovisual temporal processing development. In the experiments presented in the chapters to follow, unisensory processing abilities were likely underdeveloped in younger participants. Although some effort was made to account for this difference, investigation of the specific contribution of unisensory abilities on multisensory performance is beyond the scope of the current work. Nonetheless, future research should further explore the link between unisensory abilities and multisensory temporal processing.

Development of Multisensory Integration

Behavioral Evidence of Multisensory Development in Infancy

The development of multisensory processing has been explored in the behavioral realm in humans and two prevailing theories have emerged to explain the genesis of multisensory processing capabilities and the way in which infants determine what stimuli are related and should be integrated. The *integration view* holds that the sensory systems function independently at birth and that multisensory processing appears later in development (Birch & Lefford, 1963, 1967; Piajet, 1952). According to this theory, experience is necessary for infants and young children to learn the associations between stimuli processed across separate modalities and to subsequently bind multisensory information. Evidence for this comes from work showing an inability of infants to integrate certain forms of multisensory stimuli in early life. For example, a number of studies have shown that 4 to 8 month-old babies are unable to detect similarities in stimulus rate across sensory modalities (Humphrey & Tees, 1980; Lewkowicz, 1985, 1988a, 1992), but that 10 month-olds can successfully identify such relationships (Lewkowicz, 1988b). Moreover, 6 month-olds can detect intersensory asynchrony for dynamic and static visual stimuli, but 3 and 4 month-olds can only make such distinctions for dynamic signals (Walker-Andrews, 1986). Another established theory, the *differentiation view,* suggests that infants can integrate multisensory information from birth (Gibson, 1969; Gibson, 1966) and that this capacity is a function of the ability to extract invariant or amodal (non-modality specific) cues across the senses. For example, when watching a person clapping, the tempo characterizing both the visual motion of their hands and the auditory sound of the clapping are integrated when these amodal cues are congruent. Under this view, the neonatal system is capable of multisensory integration, but multisensory abilities are fine tuned with development. Support for an innate integrative capacity comes from research indicating that 3 week-old babies can detect interstimulus equivalence using intensity cues (Lewkowicz & Turkewitz, 1980), and two month-olds can identify congruent sights and sounds according to tempo and synchrony (Bahrick, 1988; Lewkowicz, 1992, 1996, 2000; Morrongiello, Fenwick & Chance, 1998; Morrongiello, Lasenby & Lee, 2003; Spelke, 1979). Furthermore, four to five month-old babies exhibit perception of the McGurk illusion (Burnham & Dodd, 2004; Rosenblum, Schmuckler & Johnson, 1997), although the frequency and consistency of illusion perception is reduced relative to adults (Desjardins & Werker, 2004), offering support for early integrative abilities, but continued tuning. While much work has focused on resolving whether adult-like multisensory integration is achieved via experience unifying separately processed information, or whether it is a native capacity of the newborn brain, more recently a combined model has been favored (Lickliter & Bahrick, 2004).

A hybrid model is appealing given that although infants exhibit the ability to detect some amodal relations (e.g., synchrony, tempo) and to integrate audiovisual stimuli at or shortly after birth, other capabilities emerge later (e.g., rate-based matching). Consistent with this, Lewkowicz (2000) has proposed a model in which sensitivity to temporally-based relationships emerges in a hierarchical fashion. This is supported by evidence showing that more basic features such as synchrony are detected earlier than more complex temporal structure cues (Lewkowicz, 2000). Overall findings implicate that early multisensory processing abilities becoming increasingly sophisticated with experience and growth.

Animal Models of Early Multisensory Development

Findings iterating experience dependent change in the multisensory processing abilities of infants are supported by the animal literature which has provided direct evidence of the chronology for emergence and maturation of multisensory neurons in subcortical and cortical multisensory regions of cats and rhesus monkeys. In the cat superior colliculus, a multisensory structure in the midbrain, multisensory responsive neurons emerge in the first few weeks of life (Wallace & Stein, 1997). In contrast, in more precocial species such as monkeys (Wallace & Stein, 2001), multisensory neurons are present at birth. Nevertheless, these early multisensory neurons lack integrative capabilities and mature temporal tuning, which is realized later (Wallace & Stein, 1997; Wallace & Stein, 2001). A similar but more protracted trajectory for the formation of mature multisensory neurons has been reported in areas of the cerebral cortex (Wallace, Carriere, Perrault, Vaughan & Stein, 2006).

Thus, results of physiology studies indicate that although multisensory neurons are present at or soon after birth, the establishment of mature integrating circuits occurs postnatally, implicating the possible role of sensory experience in shaping the final state of these multisensory systems. Support for the importance of sensory experience in these processes come from studies showing profound changes in multisensory integration in animals in which sensory experience is eliminated or altered (Wallace, Perrault, Hairston & Stein, 2004) (Carriere et al., 2007). Findings complement the integration and differentiation hypotheses in humans and suggest that rudimentary multisensory responses may be present at birth, but that processing matures only after a requisite period of postnatal sensory experience.

Multisensory Processing in Childhood and Adolescence

The role of experience in shaping multisensory processing has been established in infants, but developmental changes in multisensory functioning are also apparent in pre-school and school age children. Assessments of the influence of age on perception of the McGurk illusion have revealed that 4-10 year-olds experience fusions (integration) less frequently than 10-14 year-olds, 15-19 year-olds and adults, and that younger children are more likely to report the auditory token on unfused trials than their older counterparts (Massaro, 1984; McGurk & MacDonald, 1976; Tremblay et al., 2007). Moreover, multisensory facilitation of reaction times on audiovisual detection tasks is reduced in 6 to 11 year old children relative to adults, suggesting that multisensory processing immaturities continue into middle childhood (Barutchu, Crewther & Crewther, 2009; Barutchu et al., 2010). Complementary findings of developmental delays in integrative capacity have been reported in other modalities (Gori, Del Viva, Sandini & Burr, 2008; Nardini, Jones, Bedford & Braddick, 2008). For example, Gori and others (2008) found that adult-like performance on a visual-haptic task emerges between 8 and 10 years of age. Thus, while infants and young children may exhibit some forms of multisensory perception, it does not appear that the sophistication and efficiency of multisensory processing is fully realized until later. Given the paucity of studies investigating multisensory functioning in childhood, however, estimates as to the age at which mature audiovisual processing is realized are tenuous.

Multisensory Temporal Processing

Key milestones in multisensory processing development have been identified in infants and limited information is available regarding processing in children, however, the topic representing the focus of this thesis- multisensory *temporal* processing, has been largely unstudied in a developmental context. What little evidence is available suggests, perhaps not surprisingly, that the perception of multisensory temporal relationships changes with age. For example, Lewkowicz (1996) measured differences in tolerance to multisensory temporal mismatch by comparing gaze maintenance findings in infants and behavioral report in adults. He noted that thresholds for asynchrony detection in infants were more than five times greater than those of adults when the sound of a bouncing ball preceded its visual counterpart and four or more times greater when the visual stimuli came first (Lewkowicz, 1996).

While Lewkowicz's work provides preliminary evidence of developmental differences in multisensory temporal processing in infancy, there is a large gap in the literature on the role of time in modulating multisensory integration in the period between infancy and adulthood. In the studies that follow, we chose to survey performance in children and adolescents not only to fill this void, but because middle childhood has been implicated as an important transitional period for the development of mature multisensory abilities. Furthermore, the group of participants selected encompasses an age range within which all participants can successfully navigate the psychophysical measures used, enabling direct comparison of responses unlike in prior work. The assessments seek to inform us as to influence of developmental factors (e.g., system maturation, experience) on multisensory temporal processing, and its implications on overall perception and behavior.

Role of Experience in Learning Multisensory Relationships

Human behavioral studies report age-related differences in multisensory processing, and animal work provides compelling evidence of the role of experience in the development of multisensory processing, but what specific mechanisms enable these changes? Studies have suggested that humans may extract information about multisensory relationships through an implicit process known as statistical learning. This has been investigated in the context of auditory speech in infants, and is believed to partly account for the prolific gains in language ability in early life. Infants begin to show selectivity to their native language by 10-12 months of age, and lose the ability to discriminate phonetic (Werker & Tees, 1984) and visual speech (Weikum et al., 2007) contrasts of non-native languages. Moreover, infants briefly exposed to a running stream of nonsense CVs lacking pauses, stress or prosody cues, are capable of segmenting sequences based on statistical regularities learned during an exposure phase (Saffran, Aslin & Newport, 1996). Similar findings of rapid pattern detection have also been described in the visual realm where babies habituated to specific sequences of visual stimuli showed dishabituation (increased looking times) to a novel pattern as compared to the familiar one during a test phase (Kirkham, Slemmer & Johnson, 2002). Authors suggest that statistical learning spans across modalities and is therefore likely to be domain general process (Kirkham et al., 2002).

Statistical learning has also been investigated in the multisensory realm with adults. A study by Seitz and colleagues indicated that participants were able to pick up on statistical regularities (high probability sequences) to identify familiar audiovisual pairs after only an 8 minute exposure period (2007). Rapid multisensory learning effects were also reported by Kamara who showed that adults implicitly learned the relationship between a visual target and auditory cues as evidenced by reduced reaction times on a visual search task (2007).

Hence, both infants and adults are capable of learning language structure and arbitrary audiovisual relationships via identification of statistical regularities. This process may also facilitate learning of high probability (synchronous, near synchronous) multisensory temporal relationships and may promote developmental changes in multisensory temporal processing, or reduced integration of low probability (highly asynchronous) audiovisual pairings.

Specific Aims of this Thesis

Prior work has extensively characterized the multisensory temporal binding window in adults and comparison of multisensory asynchrony detection in mature subjects and infants has alluded to a developmental narrowing of this window. However, to date, no studies have investigated changes in multisensory temporal processing beyond the infant period. The work to follow represents a series of studies aimed at establishing a developmental timeline for maturation of audiovisual temporal processing with basic (non-speech) and speech stimuli. In Chapters II and III, performance is compared between developing and mature subjects on an audiovisual simultaneity judgment measure. Chapter II offers an initial survey of age-related differences in multisensory temporal processes, which served as the foundation for the work described in the chapters to follow. It established differences in audiovisual simultaneity report between 10 and 11 year-olds and adults, providing a snapshot of a single time period within the extended timeline toward realization of adult-like processing of basic stimuli. We substantially expanded on these findings in Chapter III by detailing changes in the size of the temporal binding window from early childhood to early adulthood. In this follow-up study, decreases in temporal window size were observed to increase with age, with changes in multisensory temporal processing extending into the adolescent period. Because window size is known to differ for speech and nonspeech stimuli, and given that experience is presumed to influence temporal window contraction, in Chapter IV we characterized age-related changes in multisensory integration and multisensory temporal processing to speech stimuli using the McGurk effect. The McGurk effect provides a convenient tool to index multisensory integration, and has been used to document integration of temporally synchronous and asynchronous tokens in adults, providing a comparison metric for our oldest subjects. Moreover, given the ethological validity of speech stimuli and the vast and rich experience with language from birth, it was theorized that the trajectory for window narrowing might differ from that of the low-level stimuli (i.e., rings and tone pips) presented in Chapters II and III. Surprisingly, multisensory temporal binding windows were similar in children, adolescents and adults on the McGurk task.

Despite the lack of differences in temporal processing, the likelihood of integration was reduced in younger children relative to adults at short and moderate SOAs (within the temporal window). In the General Discussion, the basis for this difference is explored as well as the implications of early experience on multisensory temporal processing.

References

Adams, R. J., & Courage, M. L. (2002). Using a single test to measure human contrast sensitivity from early childhood to maturity. *Vision Research, 42*(9), 1205-1210.

Amlot, R., Walker, R., Driver, J., & Spence, C. (2003). Multimodal visual-somatosensory integration in saccade generation. *Neuropsychologia, 41*(1), 1-15.

Bahrick, L. E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev, 59*(1), 197-209.

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Dev Sci, 12*(3), 464-473.

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. *J Exp Child Psychol, 105*(1-2), 38-50.

Birch, E. E., Gwiazda, J., Bauer, J. A., Jr., Naegele, J., & Held, R. (1983). Visual acuity and its meridional variations in children aged 7-60 months. *Vision Res, 23*(10), 1019-1024.

Birch, H. G., & Lefford, A. (1963). Intersensory Development in Children. *Monographs of the Society for Research in Child Development, 28*(5), 1-48.

Birch, H. G., & Lefford, A. (1967). Visual Differentiation, Ntersensory Integration, and Voluntary Motor Control. *Monographs of the Society for Research in Child Development, 32*(2), 1-87.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci, 21*(1), 300-304.

Buss, E., Hall, J. W., III, Grose, J. H., & Dev, M. B. (1999). Development of Adult-Like Performance in Backward, Simultaneous, and Forward Masking. *J Speech Lang Hear Res, 42*(4), 844-849.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14*(17), 2213-2218.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438.

Carriere, B. N., Royal, D. W., Perrault, T. J., Morrison, S. P., Vaughan, J. W., Stein, B. E., et al. (2007). Visual Deprivation Alters the Development of Cortical Multisensory Integration. *J Neurophysiol, 98*(5), 2858-2867.

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys, 63*(1), 126-147.

Corneil, B. D., & Munoz, D. P. (1996). The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci, 16*(24), 8193-8207.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*(4), 187-203.

Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L., & Kelso, J. A. (2007). Multisensory integration for timing engages different brain networks. *Neuroimage, 34*(2), 764-773.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Ellemberg, D., Lewis, T. L., Hong Liu, C., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*(14), 2325-2333.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J Speech Hear Res, 12*(2), 423-425.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., et al. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res, 203*(2), 381-389.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res, 147*(3), 332-343.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys, 57*(6), 802-816.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Giaschi, D., & Regan, D. (1997). Development of motion-defined figure-ground segregation in preschool and older children, using a letter-identification task. *Optom Vis Sci, 74*(9), 761-767.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr Biol, 18*(9), 694-698.

Grose, J. H., Hall, J. W., 3rd, & Gibbs, C. (1993). Temporal analysis in children. *J Speech Hear Res, 36*(2), 351-356.

Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of Spatial Contrast Sensitivity from Infancy to Adulthood: Psychophysical Data. *Optometry & Vision Science, 74*(10), 785-789.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hall, J. W., III, & Grose, J. H. (1991). Notched-Noise Measures of Frequency Selectivity in Adults and Children Using Fixed-Masker-Level and Fixed-Signal-Level Presentation. *J Speech Hear Res, 34*(3), 651-660.

Hall, J. W. I., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear and Hearing, 23*(2), 159-165.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res, 122*(2), 247-252.

Hartmann, E. E., & Banks, M. S. (1992). Temporal contrast sensitivity in human infants. *Vision Research, 32*(6), 1163-1168.

Hartmann, G. W. (1934). A Comparison of the Flicker Thresholds in Children and Adults. *Child Development, 5*(2), 122-126.

Hecht, D., Reiner, M., & Karni, A. (2008). Multisensory enhancement: gains in choice and in simple response times. *Exp Brain Res, 189*(2), 133-143.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *J Exp Psychol, 63*, 289-293.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform, 20*(1), 131-153.

Humphrey, K., & Tees, R. C. (1980). Auditory-visual coordination in infancy: Some limitations of the prefence methodology. *Bulletin of the Psychonomic Society, 16*, 213-216.

Irwin, R. J., Ball, A. K. R., Kay, N., Stillman, J. A., & Rosser, J. (1985). The Development of Auditory Temporal Acuity in Children. *Child Development, 56*(3), 614-620.

Jerger, S., & Jerger, J. (1982). Pediatric Speech Intelligibility Test: Performance-Intensity Characteristics. *Ear and Hearing, 3*(6), 325-334.

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Exp Brain Res, 174*(3), 588-594.

Jones, J. A. C. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport, 14*(8), 1129-1133.

Kawahara, J. (2007). Auditory-visual contextual cuing effect. *Percept Psychophys, 69*(8), 1399-1408.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res, 167*(4), 635-640.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-B42.

Lewkowicz, D. J. (1985). Bisensory response to temporal frequency in 4-month-old infants. *Developmental Psychology, 21*(2), 306-317.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' responses to auditory-visual compounds. *Developmental Psychology, 24*, 155-171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*(2), 172-182.

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*(3), 297-324.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform, 22*(5), 1094-1106.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull, 126*(2), 281-308.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-Modal Equivalence in Early Infancy: Auditory-Visual Intensity Matching. *Developmental Psychology, 16*(6), 597-607.

Lickliter, J. D., & Bahrick, L. E. (2004). Perceptual development and the origins of multisensory responsiveness. In G. Calvert, C. Spence & B. D. Stein (Eds.), *The handbook of multisensory processes* (pp. 643-654). Cambridge: The MIT Press.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res Cogn Brain Res, 17*(2), 447-453.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131 - 141.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev, 55*(5), 1777-1788.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am, 100*(3), 1777-1786.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: sensitivity and discrimination. *Ear Hear, 3*(6), 301-308.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America, 77*(2), 678-685.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci, 7*(10), 3215-3229.

Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res, 365*(2), 350-354.

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

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol, 75*(5), 1843-1857.

Moore, J. K., & Linthicum, F. H., Jr. (2007). The human auditory system: a timeline of development. *Int J Audiol, 46*(9), 460-478.

Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior and Development, 21*(4), 543-553.

Morrongiello, B. A., Lasenby, J., & Lee, N. (2003). Infants' learning, memory, and generalization of learning for bimodal events. *J Exp Child Psychol, 84*(1), 1-19.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol, 18*(9), 689-693.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., et al. (2007). Audiovisual Temporal Correspondence Modulates Human Multisensory Superior Temporal Sulcus Plus Primary Sensory Cortices. *J. Neurosci., 27*(42), 11431-11441.

Nozza, R. J., & Wilson, W. R. (1984). Masked and Unmasked Pure-Tone Thresholds of Infants and Adults: Development of Auditory Frequency Selectivity and Sensitivity. *J Speech Hear Res, 27*(4), 613-622.

Olsho, L. W. (1984). Infant frequency discrimination. *Infant Behavior and Development, 7*(1), 27-35.

Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., & Spetner, N. B. (1988). Pure-tone sensitivity of human infants. *J Acoust Soc Am, 84*(4), 1316-1324.

Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *J Acoust Soc Am, 82*(2), 454-464.

Pandey, P. C., Kunov, H., & Abel, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *J Aud Res, 26*(1), 27-41.

Parrish, E. E., Giaschi, D. E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Research, 45*(7), 827-837.

Piajet, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Powers, A. R., 3rd, Hevey, M., & Wallace, M. T. (Unpublished results). Neural correlates of multisensory perceptual training.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Rasengane, T. A., Allen, D., & Manny, R. E. (1997). Development of temporal contrast sensitivity in human infants. *Vision Research, 37*(13), 1747-1754.

Regal, D. M. (1981). Development of critical flicker frequency in human infants. *Vision Research, 21*(4), 549-555.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Percept Psychophys, 59*(3), 347-357.

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. *Cereb Cortex, 17*(5), 1147-1153.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science, 274*(5294), 1926-1928.

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36*(10), 1445-1453.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Res Cogn Brain Res, 14*(1), 147-152.

Sinnott, J. M., Pisoni, D. B., & Aslin, R. N. (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior and Development, 6*(1), 3-17.

Sireteanu, R., Fronius, M., & Constantinescu, D. H. (1994). The development of visual acuity in the peripheral visual field of human infants: binocular and monocular measurements. *Vision Res, 34*(12), 1659-1671.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626-636.

Spetner, N. B., & Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Dev, 61*(3), 632-652.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res, 448*(2), 355-358.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage, 44*(3), 1210-1223.

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

Teller, D. Y., McDonald, M. A., Preston, K., Sebris, S. L., & Dobson, V. (1986). ASSESSMENT OF VISUAL ACUITY IN INFANTS AND CHILDREN; THE ACUITY CARD PROCEDURE. *Developmental Medicine & Child Neurology, 28*(6), 779-789.

Teller, D. Y., & Movshon, J. A. (1986). Visual development. *Vision Res, 26*(9), 1483-1506.

Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *The Journal of the Acoustical Society of America, 98*(5), 2532-2541.

Trehub, S. E., Schneider, B. A., Morrongiello, B. A., & Thorpe, L. A. (1988). Auditory sensitivity in school-age children. *Journal of Experimental Child Psychology, 46*(2), 273-285.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PLoS One, 2*(1), e742.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Veloso, K., Hall, J. W., III, & Grose, J. H. (1990). Frequency Selectivity and Comodulation Masking Release in Adults and in 6-Year-Old Children. *J Speech Hear Res, 33*(1), 96-102.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? . *Developmental Psychology, 22*, 373-377.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Jr., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *J Neurosci, 26*(46), 11844-11849.

Wallace, M. T., Perrault, T. J., Jr., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *J Neurosci, 24*(43), 9580-9584.

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci, 17*(7), 2429-2444.

Wallace, M. T., & Stein, B. E. (2001). Sensory and Multisensory Responses in the Newborn Monkey Superior Colliculus. *J. Neurosci., 21*(22), 8886-8894.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual Language Discrimination in Infancy. *Science, 316*(5828), 1159-.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63.

Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal Resolution in Children. *Child Development, 60*(3), 611-624.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005a). Audio-visual simultaneity judgments. *Percept Psychophys, 67*(3), 531-544.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005b). Audioâ€“visual simultaneity judgments. *Perception & Psychophysics, 67*(3), 531-544.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Exp Brain Res, 152*(2), 198-210.

CHAPTER II

BINDING OF SIGHTS AND SOUNDS: AGE-RELATED CHANGES IN MULTISENSORY TEMPORAL PROCESSING*

Abstract

We live in a multisensory world in which we are continuously bombarded with sensory cues. One of the challenges the brain is faced with is deciding what information belongs together. Our ability to make assumptions about the relatedness of multisensory stimuli is partly based on their temporal and spatial relationships. Stimuli that are proximal in time and space are likely to be "bound" together by the brain and ascribed to a common external event. Using this framework we can describe multisensory processes in the context of spatial and temporal filters or "windows" that compute the probability of the relatedness of stimuli. Whereas numerous studies have examined the characteristics of these multisensory filters in adults and discrepancies in window size have been reported between infants and adults, virtually nothing is known about the developmental sequelae leading up to the mature state. To examine this, we compared the ability of 10 and 11 year olds and adults to detect audiovisual temporal asynchrony. Findings revealed striking and asymmetric age-related differences. Whereas children were able to identify asynchrony as readily as adults when visual stimuli preceded auditory cues, significant group differences were identified at moderately long stimulus onset asynchronies (150 to 350 ms) where the auditory stimulus was first. Results suggest that changes in audiovisual temporal perception extend beyond the first decade of life. In addition to furthering our understanding of basic multisensory developmental processes, these findings have implications on disorders (e.g., autism, dyslexia) in which emerging evidence suggests alterations in multisensory temporal function.

Keywords: intersensory, auditory, visual, development, simultaneity, asynchrony

---

* The contents of this chapter were submitted as:
Hillock, A.R., Powers, A.R., and Wallace, M.T. Binding of sights and sounds: Age-related changes in multisensory temporal processing.

Introduction

Many of our everyday experiences are multisensory. For example, in a typical communicative exchange, we hear the words that are spoken and see the corresponding visual information in the form of an individual's lips moving. To make sense of the wealth of sensory information available at any given moment, our brains have evolved specialized mechanisms to extract meaningful information both within and across the different sensory systems. For multisensory processes, two of the most salient pieces of information used to determine the relatedness of objects or events are their spatial and temporal proximity, and numerous studies in adults have focused on defining how manipulations of these relations alter the magnitude of multisensory interactions. Despite this wealth of data, surprisingly little is known about how these processes mature during postnatal life. Based on the premise that judgments regarding the interrelatedness of multisensory stimuli are modified with age, this study explores the influence of development on multisensory processing by contrasting audiovisual temporal asynchrony detection abilities in younger and older participants.

Temporal Aspects of Multisensory Processing

The benefits of the combined use of information from several senses have been revealed in numerous studies, and include enhancements in signal detection, speeded motor responses, and improved speech in noise performance (Frens, Van Opstal, & Van der Willigen, 1995; Grant & Seitz, 2000; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Lovelace, Stein, & Wallace, 2003; Sumby & Pollack, 1954). In addition, a host of psychophysical illusions reveal the continual and compelling interplay between the senses. For example, pairing distinct and discordant auditory and visual speech cues (e.g. an auditory /ba/ with a visual /ga/) can result in report of an intermediary and novel percept (e.g. /tha/ or /da/) (McGurk & MacDonald, 1976). Illusory percepts can also be generated with highly reduced multisensory stimuli, as evidenced by the fact that the presentation of a single visual flash accompanied by two tone pips can result in strong report of multiple flashes (Shams, Kamitani, & Shimojo, 2000).

Where examined, these behavioral benefits and perceptual illusions have been shown to be critically dependent on the temporal and spatial structure of the paired stimuli, with the strength of multisensory interactions declining as a function of increasing spatial and/or temporal disparity. In the temporal realm, numerous studies have suggested the presence of a temporal “window” of multisensory integration, or a range of interstimulus intervals over which multisensory stimuli are highly likely to be bound into a single perceptual event (Dixon & Spitz, 1980; Shams, Kamitani, & Shimojo, 2002; van Wassenhove, Grant, & Poeppel, 2007). The boundaries of these temporal binding processes have been delineated by quantifying the perseverance and magnitude of multisensory effects (e.g. speeded motor reaction times, psychophysical illusions, reports of simultaneity) as the time interval between the presentation of the constituent multisensory stimuli is lengthened (Colonius & Diederich, 2004; Dixon & Spitz, 1980; Koppen & Spence, 2007; Munhall, Gribble, Sacco, & Ward, 1996; Shams, et al., 2002; van Wassenhove, et al., 2007).

In the search for the neural correlates for these multisensory behavioral and perceptual phenomena, human electrophysiological and imaging studies have revealed a similar temporal dependency. Synchronous presentation of auditory and visual speech produces decreases in the latency of early cortical auditory evoked potentials (van Wassenhove, Grant, & Poeppel, 2005). Furthermore, combined audiovisual stimuli produce greater gamma band oscillatory activity when presented at smaller audiovisual stimulus onset asynchronies (SOAs) (Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007), which is suggested to be reflective of multisensory binding (Schneider, Debener, Oostenveld, & Engel, 2008; Senkowski, Schneider, Tandler, & Engel, 2009). Imaging studies have identified a distributed network of cortical and subcortical regions involved in multisensory integration whose activation profiles reveal a strong temporal dependency (Bushara, Grafman, & Hallett, 2001; Dhamala, Assisi, Jirsa, Steinberg, & Kelso, 2007; Noesselt, et al., 2007; Powers, Hevey, & Wallace, Unpublished results). For example, blood oxygenation level dependent (BOLD) signals, an indirect measure of neuronal activity revealed in fMRI studies, have been found to be increased in the superior temporal sulcus (STS) and auditory and visual cortices in response to coincident audiovisual stimuli (Noesselt, et al., 2007). Conversely, BOLD decreases have been observed in these areas in response to asynchronous stimulus pairs (Noesselt, et al., 2007).

Age Effects on Multisensory Task Performance

While the bulk of human multisensory research has focused on adults, the development of various aspects of multisensory integration has been examined in human infants, largely using paradigms that track the duration of gaze maintenance (i.e., preferential looking). Behavioral work has indicated that infants as young as four months of age show the ability to detect tempo and synchrony, and that temporal invariants (amodal cues- available to both auditory and visual senses) assist in discerning what stimuli are produced by a unitary event (Bahrick, 1983; Bahrick, 1987, 1988; Lewkowicz, 1986, 1992, 1996, 2000; Spelke, 1979). Furthermore, Lewkowicz (1996) has shown that when compared to adults, infants have a larger temporal “window” for binding asynchronous visual and auditory stimuli, suggesting that they perceptually fuse temporally disparate multisensory stimulus pairs that are not fused in adults.

While no studies have examined multisensory performance between older children and adults on temporally-based tasks, work has reported immature audiovisual and visual-haptic processing abilities in pre-adolescents and adolescents relative to adults on a variety of non-temporally-based tasks. A study examining the McGurk illusion in children and adults found that when responses were dominated by a single modality (i.e., unfused trials), children relied more heavily on the auditory input while adults placed greater weight on the visual cue (McGurk & MacDonald, 1976). Similar differences were also reported on an audiovisual syllable identification task with children 4-9 and 6-9 years of age and adults (18-38 years) (Massaro, 1984). Performance differences have also been noted in children 5 to 10 years of age and adults on a visual and haptic (active touch) size discrimination task (Gori, Del Viva, Sandini, & Burr, 2008). Whereas adults and older children were found to weight visual information more heavily in making size estimations, the judgments of the youngest children were more heavily influenced by the haptic information. Data suggested that middle childhood (i.e., 8-10 years) may represent an important transitional period for the maturation of adult-like multisensory processing.

Multisensory Temporal Processing in Children: Rationale for the Current Study

As highlighted, research has identified age-related differences in multisensory abilities and has established the concept of a plastic multisensory temporal binding window that changes with ontogeny. However, no studies have systematically documented developmental differences in the temporal aspects of audiovisual integration between children and adults. The goal of the current study was to better characterize multisensory temporal function in children and adults using an audiovisual simultaneity judgment task previously used by our group to assess the multisensory temporal binding window in adults (as well as its malleability, see Powers et al. 2009). This measure enables us to directly compare multisensory temporal function in children and adults.

Methods

Subjects

Typically developing children (n=13; 11 males; mean age= 10.7 years) and adults (n=14; 6 males; mean age= 26.6 years) were recruited to participate in the study. All participants and parents/guardians of minors were consented and assented prior to study participation in accordance with an approved protocol of the Vanderbilt Institutional Review Board (IRB). All subjects had normal hearing (pure tone thresholds less than 25dB HL at octave frequencies from 250 to 8,000 Hz) and good visual acuity with or without correction (Snellen criterion of 20/25 – 2 [2 or fewer errors per eye]) and average or above-average intelligence. All children completed the Kaufmann Brief Intelligence Test, second edition (K-BIT II), which provides an estimate of intellectual ability. No formal intelligence screening was completed on adults. The adult group was comprised of college educated individuals and undergraduate students at Vanderbilt University with no history of learning difficulties (per self report). Hearing and vision screenings were completed at the start of the session and intelligence testing followed completion of the multisensory task.

Stimuli and Experimental Design

A point of subjective simultaneity (PSS) judgment task (adapted from Fujisaki et al., 2004) was administered to each participant twice (i.e., assessment 1 and assessment 2). Subjects were seated in a quiet, dimly lit room approximately 48 cm from a high refresh-rate computer monitor (ViewSonic E70fB, 120 Hz). A white crosshair fixation marker (1 cm x 1 cm) was situated in the center of a black background on the computer screen for the duration of the experiment. Auditory (8 ms duration, 1800Hz tone burst, 99.2 peak dB SPL [unweighted]) and visual (8 ms duration, white ring flash subtending 15° of space, outer diameter = 12.4 cm, inner diameter = 6.0 cm, area = 369.8 cm^2) stimulus pairs were presented in a randomly interleaved fashion at the following visual-auditory SOAs: 0, +/- 50, +/- 100, +/- 150, +/- 200, +/- 250, +/- 350 and +/- 450 ms (Figure 2.1).

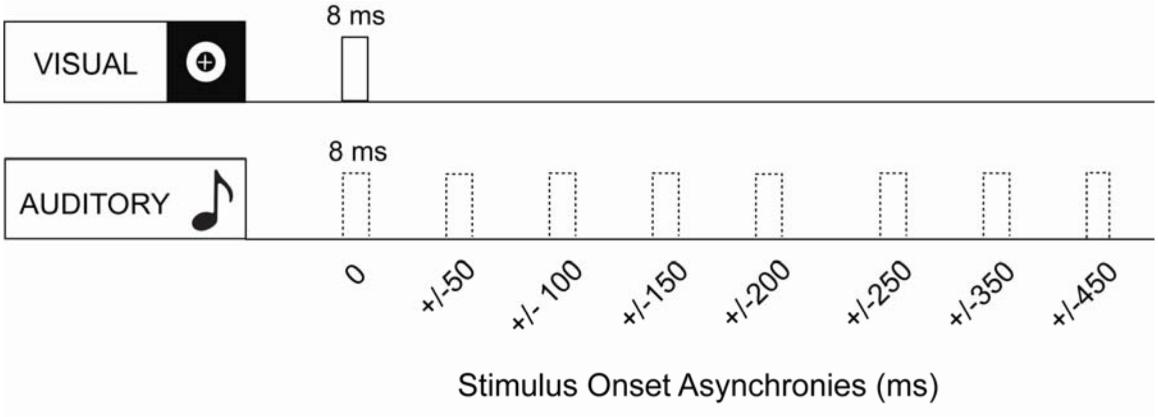


Figure 2.1. Simultaneity Judgment Paradigm. *Profile of temporal relationship between stimuli used in assessments.*

The auditory stimulus was presented via Etymotic Research ER-3A insert earphones and auditory and visual stimulus duration and SOAs were verified externally with an oscilloscope. Whereas positive values represent visual stimuli leading auditory stimuli, negative values represent the opposite. A total of 330 responses were collected during each assessment (22 samples x 15 SOAs). Each new trial was initiated 1 second after the participant logged his/her response to the previous trial. Total test time for the multisensory task was approximately 25 minutes, including breaks. Stimulus delivery and data logging was controlled by E-prime 2.0 (2.0.1.109).

Instructions were read aloud for both groups and the task was initiated by the experimenter (see Appendices 2.1 and 2.2 for adult and child instructions). For the children, the task objectives were placed in the context of a story that distinguished the auditory-visual communication styles of boy and girl lightning bugs. Following the story, children answered a circumscribed list of questions to ensure proper understanding of the task and the capacity to distinguish between the bug images (Appendix 2.3). Behavioral judgments (i.e., simultaneous or asynchronous) were recorded by pressing buttons labeled with numbers (1 or 4, adults) or lightning bug images (blue [male] or red [female], children) on a response box (Psychology Software Tools Response Box Model 200A). Bug images had characteristics other than color to aid in distinguishing their gender (as screenings did not include testing for color vision deficiency). Responses were counterbalanced across participants such that the buttons associated with simultaneity and non-simultaneity were reversed for half of the subjects. A follow-up experiment in which a novel group of adults (n=8) were instructed using the story-based technique employed with the children revealed no significant effect of instruction method on responses compared to those of 8 randomly selected adults from our experimental group.

Data Analysis and Temporal Window Derivation

Differences in window size and simultaneity judgment at sampled SOAs were assessed between groups and across assessments using independent samples *t*-tests and repeated-measures analysis of variance (rmANOVA). To correct for dependence among the repeated measures within subjects, Greenhouse-Geisser corrections were performed. Because no significant effect of session was observed for measures of window size (Supplement 2.1), the data were collapsed across assessments. However, for analysis of within-session effects (e.g., effects resulting from fatigue or task-learning), responses were divided between the first and second halves of the assessments.

The mean probability of simultaneity report (i.e., number of simultaneous responses as a function of total responses) was calculated for responses at each SOA for all participants. Points were fit to create a distribution which served as the basis for determining the size of the temporal binding window. Sigmoids were generated to the discretely sampled points from each half of the distribution (left: -450 to 0 ms, right: 0 to +450 ms) for all subjects. Distributions were produced from a two-by-two matrix comprised of interpolated y values (probability of simultaneity report) at x values (time points) ranging from -600 to +600 ms in 0.1 ms increments. The criterion used to characterize the size of the multisensory temporal binding "window" for each subject was the width of this distribution at 3/4 maximum.

Results

Defining Group Differences in Window Size

The mean multisensory temporal binding window for the 10 and 11 year old children (413 ms) was found to be significantly wider than that of the adults (299 ms) ($t(25)= 3.945$, $p= 0.001$; Table 2.1 and Figure 2.2). However, several children had window sizes comparable to the adult adults. In general, children were found to have more symmetric distributions when compared with adults, as illustrated by the comparison in two representative subjects shown in Figure 2.3.

Table 2.1. Group differences in window size. *Summary table of differences in window size between children and adults (assessments averaged).*

	Mean WS (ms)	SEM	*t*	df	p
Child Group (n=13)	412.762	23.155	3.945	25	*p=0.001
Adult Group (n=14)	298.929	17.642			

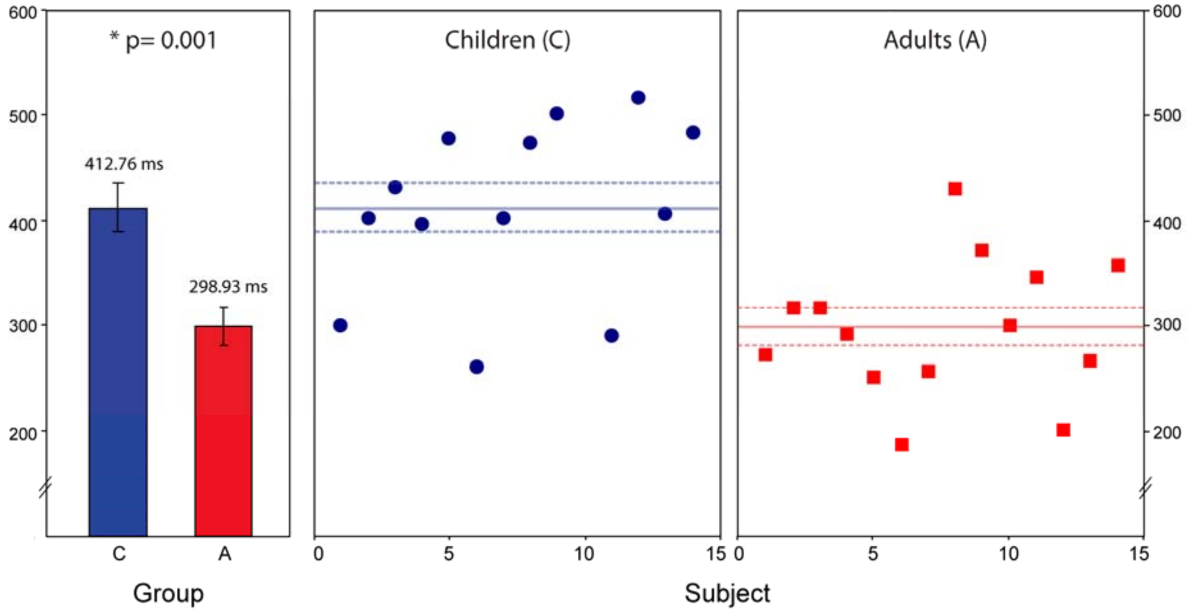


Figure 2.2. Mean window size in children is larger than that in adults. *Left:* Bar graph displays group difference in multisensory temporal window size (i.e., distribution width at 3/4 maximum) for children (C, blue) and adults (A, red). Error bars represent +/- one standard error of the mean (SEM). *Middle:* Scatterplot of individual window sizes for each child (blue circle) and adult (red square). Solid lines represent mean values and dotted lines denote the area encompassing +/- 1 SEM.

mean, more than a third of the children had windows in excess of 450 ms, a value never seen in

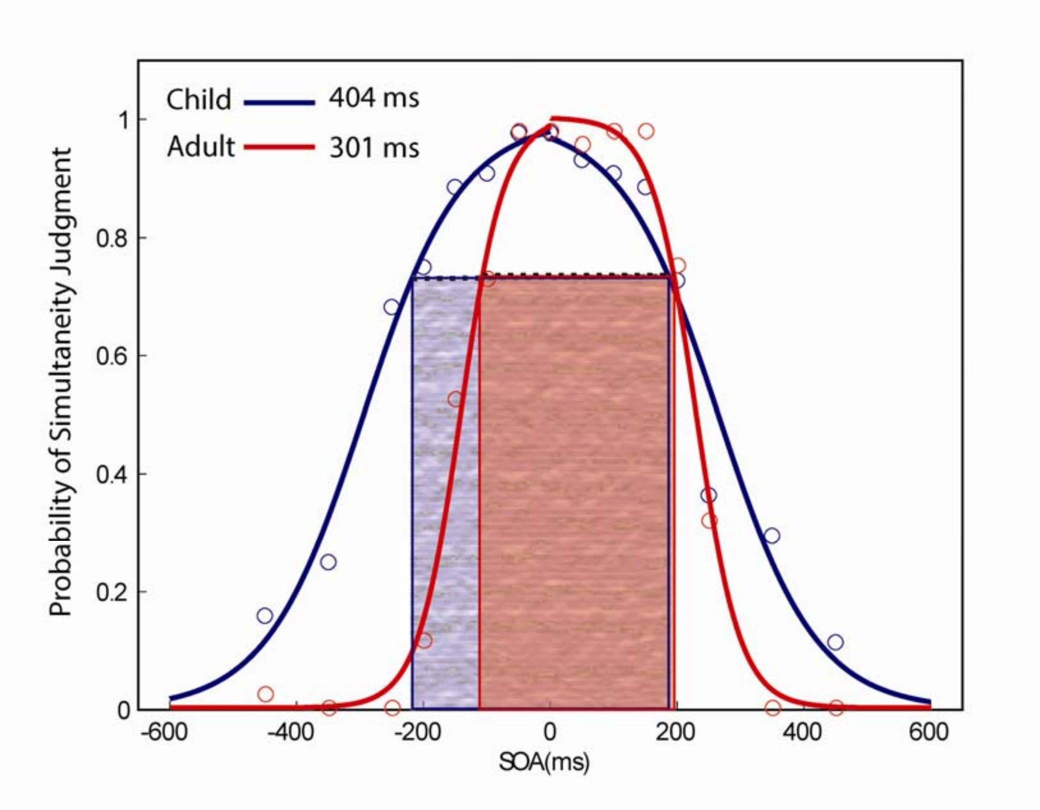


Figure 2.3. Data in representative subjects shows differences in overall window size and response distribution symmetry. *Overall window size for the child subject is more than100 ms greater than that of the adult. Note that the discrepancy in window size between subjects is primarily comprised of differences in the left side of the distribution.*

Defining Group Differences in the Probability of Simultaneity Judgment

Although the comparison of mean window size provided a global measure of group differences in the multisensory temporal binding window, additional analyses were performed to further delineate the temporal structure and nature of the age-related differences in participants' reports of simultaneity/asynchrony. A repeated-measures ANOVA with a within-subjects factor of SOA and a between-subject factor of group revealed a significant main effect of group and SOA and an SOA by group interaction, $p< 0.01$, all tests (Table 2.2). These results indicate group differences in simultaneity judgment that differ across SOA conditions, or distinctions between groups at select SOAs.

Table 2.2. Main effects and interaction effect. *Summary table of significant effects from a repeated-measures ANOVA using a repeated-measures factor of SOA and a between subjects factor of group (assessments averaged).*

Factor	F	df factor	df error	p
SOA	154.691	3.502	87.548	* 0.000
Group	15.120	1	25	* 0.001
SOA * Group	4.094	3.502	87.548	* 0.006

Follow-up contrasts (i.e., independent samples *t*-tests) were performed to identify which SOAs contributed to these group differences. A Bonferonni correction was applied to reduce the risk of elevation of familywise error due to the number of comparisons (i.e., 15), resulting in a more stringent criterion value for significance (p< 0.0033). Equal variance between groups was assumed for all comparisons unless otherwise noted based on results from Levene's test for equality of variances. Significant differences in the likelihood of simultaneity report were observed for the following conditions: -350, -250, -200, -150 ms (Figure 2.4, Table 2.3) – moderate to moderately long SOAs in which the onset of the auditory stimulus preceded the onset of the visual stimulus. A marginally significant difference was noted between groups at the most extreme auditory leading visual SOA (-450 ms), p=0.004. In contrast, no differences in simultaneity report were observed in the objectively simultaneous condition, at short stimulus onset asynchronies (e.g., 50ms, 100ms) in which the auditory stimulus was leading, or in any of the conditions in which the visual stimulus preceded the auditory stimulus.

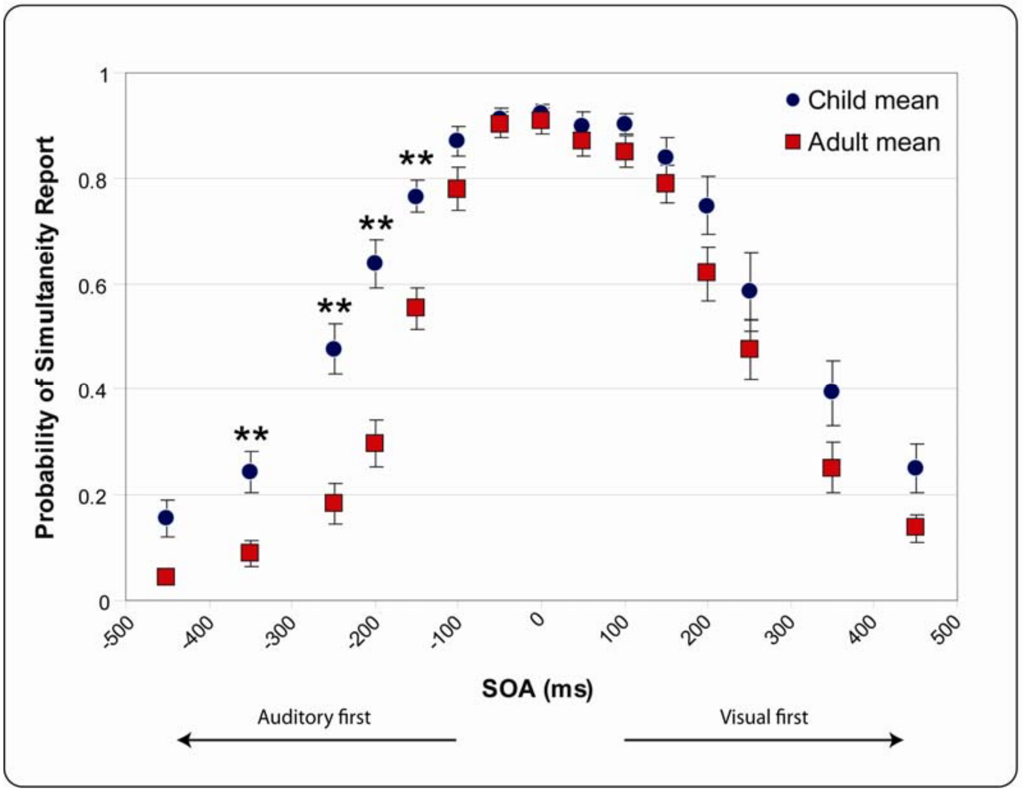


Figure 2.4. Grand averaged group data distributions. *Children were significantly more likely to report trials as simultaneous at moderate and long SOAs in which the auditory stimulus preceded the ring flash. Error bars represent +/- 1 SEM,** sig Bonferroni $p<0.0033$.*

Table 2.3 Specific SOAs with group differences. *Summary table of SOAs showing significant group differences in probability of simultaneity report (assessments averaged).*

		Mean	SEM	Min	Max	*t*	df	*p*
-350 ms	Child (n=13)	0.243	0.039	0.045	0.409	3.404	20.143	*0.003
	Adult (n=14)	0.086	0.024	0.000	0.273			
-250 ms	Child (n=13)	0.476	0.048	0.159	0.682	4.797	25	* 0.000
	Adult (n=14)	0.183	0.039	0.000	0.477			
-200 ms	Child (n=13)	0.636	0.046	0.318	0.886	5.311	25	* 0.000
	Adult (n=14)	0.297	0.045	0.114	0.591			
-150 ms	Child (n=13)	0.766	0.031	0.591	0.978	4.208	25	* 0.000
	Adult (n=14)	0.552	0.040	0.318	0.795			

Equal variances not assumed for the A350V condition only.

Summary of Between and Within Assessment Group Effects

In an effort to assess possible fatigue and procedural or perceptual learning effects, subjects' responses were compared across the two assessments (Figure 2.5). These pseudocolor plots serve to reinforce the difference between children and adults in the size of the temporal window (C – A, bottom contrast – note the difference being more pronounced on the left side of the distributions), as well as showing equivalent performance between the two assessments (A1 – A2, right contrast).

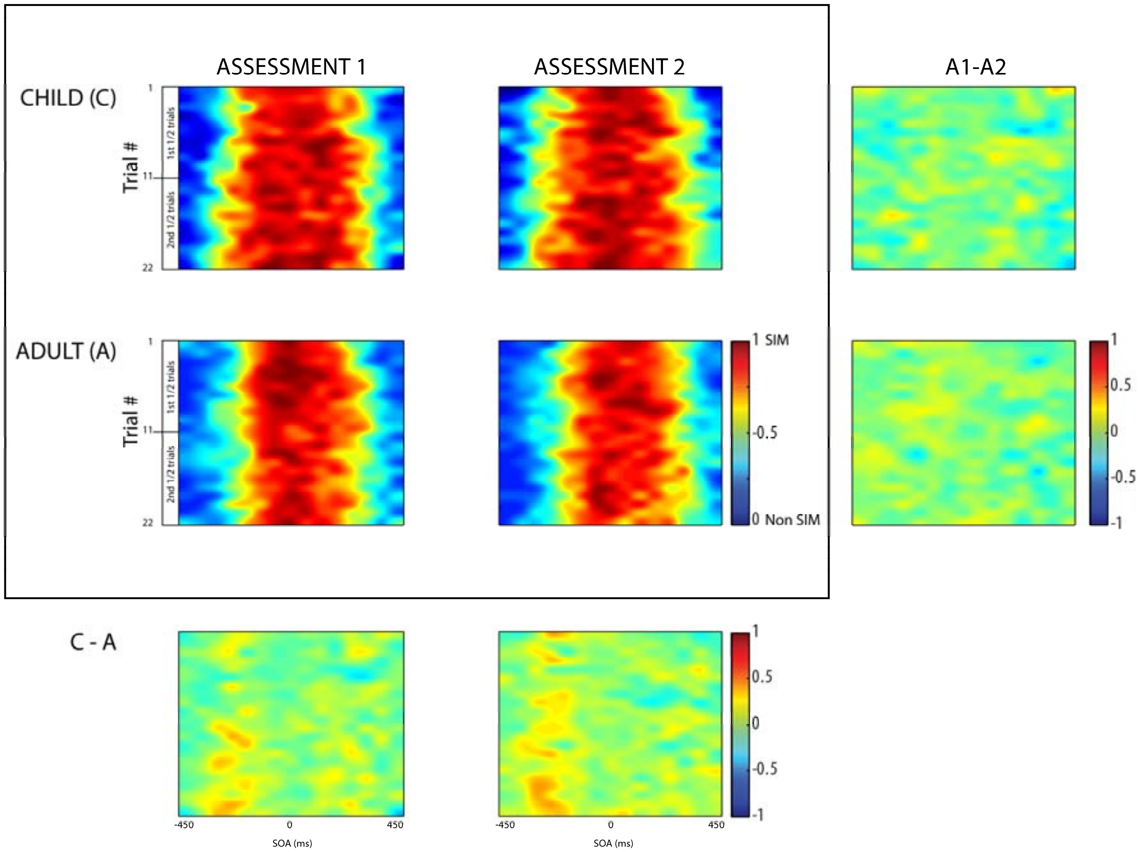


Figure 2.5. Comparison of within-session performance between groups and within groups. *Smoothed pseudocolor plots depicting mean probability of simultaneity judgment for each group on consecutive trials (top left) where warmer and cooler colors indicate a higher probability of simultaneity and asynchrony report, respectively. Trials are aligned from first (top) to last (bottom). Contrast plots reveal differences in behavioral report between groups (bottom) and across assessments (top right). For these contrast plots warmer and cooler colors represent positive and negative remainders, respectively; green is neutral or no change. Note the consistency of responses across assessments for both groups (top right) and the group difference in the simultaneity report at negative SOAs (bottom).*

Despite the lack of differences across assessments for either group, additional analyses were performed to assay more rapid within-session changes. In this analysis, windows were derived from averaged responses collected during the first half of assessments 1 and 2 and compared to those collected in the latter half of the assessments. In adults, no significant within assessment effects were observed ($p > 0.05$). In contrast, windows derived from children's' responses on the last half of trials were larger than those from the first half of responses, $t(12) = -2.496$, $p = 0.028$

(Table 2.4). Mean window size in children differed by 79 ms, whereas in adults this difference was only 12 ms.

Table 2.4. Group differences within assessments. *Summary table showing significant differences in children's' windows derived from averaged responses during the first half of the assessments (i.e., A12_first) relative to the latter part (i.e., A12_last) of the assessments.*

		Mean	SEM	*t*	df	p
A12_first	Child (n=13)	370.046	23.931	2.972	12	* 0.028
A12_last	Child (n=13)	449.462	34.094			
A12_first	Adult (n=14)	291.100	17.012	.788	13	0.445
A12_first	Adult (n=14)	279.057	23.791			

To determine whether the widened distributions exhibited by children during the latter part of each experiment were driving the group differences, window sizes derived from the earlier and later trials were compared across groups. Results of independent samples *t* tests revealed significant group differences in both the beginning and end of assessments 1 and 2 (Table 2.5). While these findings suggest that children fatigued more quickly than adults, illustrated by window widening in the latter half of the assessments, the enlargement of window size was still evident in the earliest trials. Thus, while group differences may be more pronounced in the latter portion of the assessments, age effects are apparent throughout the experimental task suggesting that differential fatigue and/or learning effects cannot fully account for group differences.

Table 2.5. Within group assay of within assessment changes. *Summary table of group differences in windows derived from averaged responses from the first half of trials in both assessments and then the latter half of trials.*

		Mean	SEM	*t*	df	p
A12_first	Child (n=13)	370.046	23.931	2.719	25	* 0.012
	Adult (n=14)	291.100	17.012			
A2_last	Child (n=13)	449.462	34.094	4.148	25	* 0.000
	Adult (n=14)	279.057	23.791			

Discussion

The current study used an audiovisual point of subjective simultaneity task to reveal age-related differences in the temporal window of multisensory integration. Most surprisingly, the results illustrate that even in children ages 10 and 11 the multisensory temporal binding window is far from mature. As discussed below, such a result has important implications in a variety of domains.

The broader temporal binding window in children appears to be driven largely by significant differences in the maturation of the left side of the measured response distributions. Children were significantly more likely to report simultaneity under conditions in which the auditory stimulus led the visual stimulus when compared with adults (but not the converse). In essence, the results show greater symmetry in the overall response distributions in children when compared with adults, and suggest that the narrowing of the left side of the distribution must take place after ages 10 -11. The reason for this asymmetrical developmental effect is unknown, but may be related to the physical characteristics of audiovisual stimuli in a naturalistic environment. Under realistic circumstances, for a given stimulus event that generates multisensory energies, visual signals will always lead the associated auditory signals (a result of the fact that visual signals travel at the speed of light whereas auditory signals are delayed in a distance dependent manner). Consequently, extensive experience with audiovisual stimuli during development may drive the meaningful side of the distribution (positive SOAs) to mature more quickly. This still leaves open the question as to why the right side matures so late (or changes at all given its lack of real world relevance).

Mechanistically, the enlarged temporal binding window must reflect anatomical and/or physiological differences in the circuits that serve to bind together multisensory cues. Several neuroimaging studies in adults have begun to identify important nodes in a network subserving audiovisual simultaneity perception, These include the insula (Calvert, Hansen, Iversen, & Brammer, 2001), inferior parietal lobule (Calvert, et al., 2001; Dhamala, et al., 2007), superior colliculus (Calvert, et al., 2001; Dhamala, et al., 2007), posterior superior temporal sulcus (Calvert, Campbell, & Brammer, 2000; Calvert, et al., 2001; Dhamala, et al., 2007), and unisensory cortices (Noesselt, et al., 2007). These latter two have also been shown to exhibit altered BOLD activity and effective connectivity after training on a simultaneity judgment task (Powers, et al., Unpublished results). Incomplete myelination of tracts incorporated in the networks involved in multisensory processing may contribute to variable neural propagation times, which could partly account for enlarged windows seen in infants (Lewkowicz, 1996). In addition, maturational differences are likely to be a result, at least in part, to the slow functional maturation of cortical networks involved in these binding processes. Neuroimaging studies have revealed a surprisingly extended time course for the complete maturation of cortical function. Linear increases in white matter volume have been reported throughout the brain from early childhood through the second decade of life (4-20 years of age). Gray matter volume reportedly peaks during adolescence or early adulthood (depending on brain region) and declines thereafter with higher order association cortices (i.e., prefrontal cortex, superior temporal cortex) maturing up to and potentially beyond 20 years of age (Giedd, et al., 1999; Gogtay, et al., 2004; Pfefferbaum, et al., 1994).

Another important determinant in the development of multisensory functioning is the acquisition of increasing amounts of sensory experience. Prior work has suggested a hierarchy of multisensory temporal processing abilities given that different audiovisual capabilities mature at different rates. Thus, while detection of audiovisual synchrony has been observed in infants as young as 4 months, identification of duration-based correspondences (between synchronous audiovisual stimuli) and sensitivity to rate-based manipulations does not emerge until 6 months and 10 months of age, respectively (Lewkowicz & Lickliter, 1994). It has also been speculated that shifts in sensory dominance may contribute to changes in infants' responsiveness to intersensory incongruence. Discrepancies between the onset of functional hearing (third trimester) and vision (birth) can lead to increased reliance and attention to the more developmental mature sense, audition, in early life. To that end, while younger (6 month old) infants are exclusively able to detect audiovisual intersensory incongruence when the auditory stimulus or both the auditory and visual stimuli are manipulated, older (10 month old) babies with more visual experience show detection of auditory, visual or audiovisual stimulus manipulations (Lewkowicz & Lickliter, 1994). Thus, as infants gain more experience interacting with the sensory world, their intersensory matching abilities and strategies are modified.

Behavioral studies in pre-adolescent children and adults also suggest an age and experience dependent change in multisensory processing and report a link between performance improvements and optimization of statistical cue weighting. It has been theorized that different sensory inputs are weighted on the basis of their relative reliability when combined within the nervous system (Alais & Burr, 2004; Ernst & Banks, 2002; Helbig & Ernst, 2007; Wozny, Beierholm, & Shams, 2008). Because sensory systems mature at different rates, the relative reliability of the information coded by each system shifts, subsequently altering the weights attributed to respective cues. Interestingly, and germane to the current study, recent work has suggested that this statistical optimization of multisensory integration is not realized until middle childhood (Gori et al., 2008; Nardini et al., 2008). If cue reliability weightings are calculated over a specific temporal interval, then the delayed maturation of the multisensory temporal binding window may interfere with the optimality of this process. Future work that relates the developmental time course of multisensory temporal function and statistical optimality will shed light on this possibility.

Research from this and other laboratories has indicated that multisensory processing may be altered in developmental disabilities such as dyslexia and autism (Hairston, Burdette, Flowers, Wood, & Wallace, 2005; Hari & Renvall, 2001; Laasonen, Service, & Virsu, 2001, 2002; Laasonen, Tomma-Halme, Lahti-Nuuttila, Service, & Virsu, 2000; Lovaas, Schreibman, Koegel, & Rehm, 1971; Mongillo, et al., 2008; Smith & Bennetto, 2007; Virsu, Lahti-Nuuttila, & Laasonen, 2003; Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004). The goal of the current study was to document age-related temporal multisensory processing differences between adults and typically-developing 10 and 11 year old children. Future research will expand this to additional age groups with the ultimate goal of creating a developmental trajectory for normative multisensory temporal processing. Identification of these benchmarks and chronology will be of tremendous use in the screening and treatment of these developmental disabilities. For example, recent work has found that perceptual training on a simultaneity judgement task identical to that employed here can narrow the multisensory temporal window in typical adults (Powers, Hillock, & Wallace, 2009). Application of such training methods to impaired populations could hold great promise in the remediation of multisensory deficits, and by extension those higher order processes dependent on the faithful binding of multisensory cues.

Acknowledgments

We thank Drs. Linda Hood, Wesley Grantham, Alexandra Key and David Royal for their intellectual contributions as well as Dr. Lynnette Henderson for her assistance with subject recruitment.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development, 6*, 429-451.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior & Development, 10*, 387-416.

Bahrick, L. E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev, 59*(1), 197-209.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci, 21*(1), 300-304.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol, 10*(11), 649-657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci, 16*(6), 1000-1009.

Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L., & Kelso, J. A. (2007). Multisensory integration for timing engages different brain networks. *Neuroimage, 34*(2), 764-773.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys, 57*(6), 802-816.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci, 2*(10), 861-863.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A, 101*(21), 8174-8179.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr Biol, 18*(9), 694-698.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am, 108*(3 Pt 1), 1197-1208.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn Sci, 5*(12), 525-532.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Exp Brain Res, 179*(4), 595-606.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform, 20*(1), 131-153.

Koppen, C., & Spence, C. (2007). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Res, 1186*, 224-232.

Laasonen, M., Service, E., & Virsu, V. (2001). Temporal order and processing acuity of visual, auditory, and tactile perception in developmentally dyslexic young adults. *Cogn Affect Behav Neurosci, 1*(4), 394-410.

Laasonen, M., Service, E., & Virsu, V. (2002). Crossmodal temporal order and processing acuity in developmentally dyslexic young adults. *Brain Lang, 80*(3), 340-354.

Laasonen, M., Tomma-Halme, J., Lahti-Nuuttila, P., Service, E., & Virsu, V. (2000). Rate of information segregation in developmentally dyslexic children. *Brain Lang, 75*(1), 66-81.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development, 9*(3), 335-353.

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*(3), 297-324.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*(5), 1094-1106.

Lewkowicz, D. J. (2000). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development, 71*(5), 1241-1257.

Lewkowicz, D. J., & Lickliter, R. (Eds.). (1994). *The development of intersensory perception: comparative perspectives.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory input. *J Abnorm Psychol, 77*(3), 211-222.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res Cogn Brain Res, 17*(2), 447-453.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev, 55*(5), 1777-1788.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *J Autism Dev Disord, 38*(7), 1349-1358.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., et al. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci, 27*(42), 11431-11441.

Pfefferbaum, A., Mathalon, D. H., Sullivan, E. V., Rawles, J. M., Zipursky, R. B., & Lim, K. O. (1994). A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Arch Neurol, 51*(9), 874-887.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Powers, A. R., Hevey, M., & Wallace, M. T. (Unpublished results). Neural correlates of multisensory perceptual training.

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage, 42*(3), 1244-1254.

Senkowski, D., Schneider, T. R., Tandler, F., & Engel, A. K. (2009). Gamma-band activity reflects multisensory matching in working memory. *Exp Brain Res, 198*(2-3), 363-372.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45*(3), 561-571.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Res Cogn Brain Res, 14*(1), 147-152.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *J Child Psychol Psychiatry, 48*(8), 813-821.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626-636.

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

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

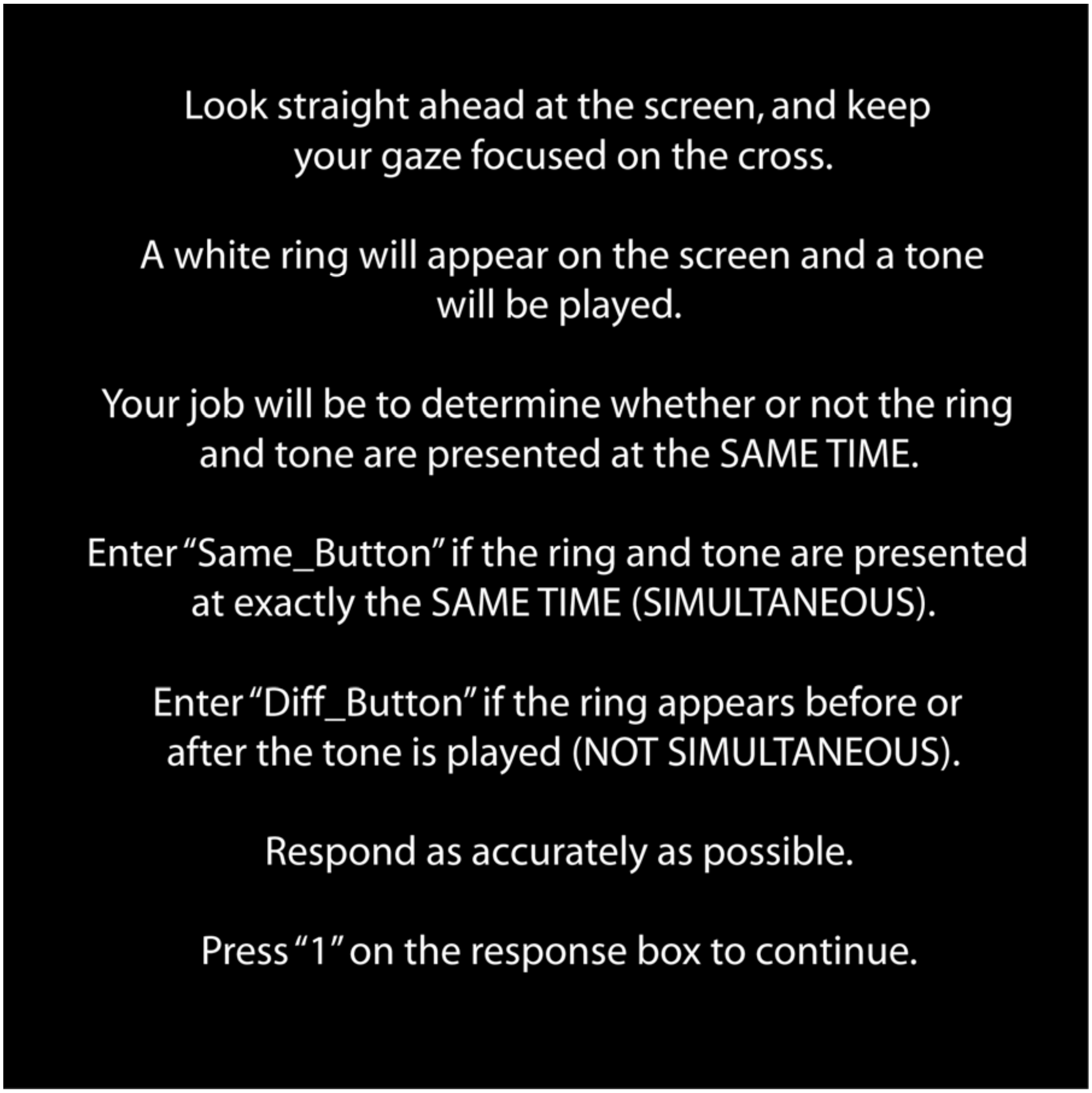
Virsu, V., Lahti-Nuuttila, P., & Laasonen, M. (2003). Crossmodal temporal processing acuity impairment aggravates with age in developmental dyslexia. *Neurosci Lett, 336*(3), 151-154.

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Res Dev Disabil, 25*(6), 559-575.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *J Vis, 8*(3), 24 21-11.

Appendix

Appendix 2.1. PSS task instructions for adults.



Look straight ahead at the screen, and keep your gaze focused on the cross.

A white ring will appear on the screen and a tone will be played.

Your job will be to determine whether or not the ring and tone are presented at the SAME TIME.

Enter "Same_Button" if the ring and tone are presented at exactly the SAME TIME (SIMULTANEOUS).

Enter "Diff_Button" if the ring appears before or after the tone is played (NOT SIMULTANEOUS).

Respond as accurately as possible.

Press "1" on the response box to continue.

Appendix 2.2. Child PSS (CPSS) story containing instructions for 10 and 11 year olds.

CPSS STORY

The Great Bug Escape

NASHVILLE ZOO AT GRASSMERE

Earlier today at the Nashville Zoo the boy and girl lightning bugs got out of their cages and got mixed up. They need to be returned to their separate homes by tomorrow morning. 1

It is getting late and has become dark outside. You can't see the bugs. The only way you can tell the lightning bugs apart is by their chirps and flashes. The lightning bugs' tails look like this circle…

2

The boy lightning bugs chirp and flash their tail at exactly the **SAME** time. 3

The girl lightning bugs chirp BEFORE or AFTER they flash their tails. The girl bugs NEVER chirp and flash at the same time. 4

Your job is to help us sort the lightning bugs. When you hear the bug chirp and see a flash at the **SAME** time, press the button for the **boy** bug.

When the chirp and flash **do NOT** happen at exactly the same time, press the button for the **girl** bug.

Do your best! Thanks for your help! 5

The End! 6

Appendix 2.3. CPSS story pre-test questions.

1) Referring to response box:

Point to the girl bug?

Point to the boy bug?

2) Which bug chirps & flashes at exactly the SAME time?

3) Which bug chirps BEFORE or AFTER it flashes?

Supplementary Data

Supplement 2.1. *Table of rmANOVA using window size data and between-subjects factor of group (child, adult) and within-subjects factor of assessment (1, 2). A main effect of group was observed, but the effect of assessment and assessment by group interaction were not significant. (Justification for combination of assessments for analysis.)*

Factor	F	df factor	df error	P
Assessment (1,2)	2.522	1	25	0.125
Group (C,A)	11.831	1	25	* 0.002
Assessment * Group	2.606	1	25	0.119

CHAPTER III

CHANGES IN THE MULTISENSORY TEMPORAL BINDING WINDOW PERSIST INTO ADOLESCENCE

Abstract

We live in a world rich in sensory information. Consequently, one of the challenges that the brain is tasked with is deciphering which cues from the various sensory modalities belong together and which should be segregated. Such determinations regarding the relatedness of sensory information are likely to be based, at least in part, on the spatial and temporal relationships of the stimuli. For example, in the temporal realm, stimuli occurring at or near the same moment in time are highly likely to be associated with one another, whereas those that are sufficiently separated in time are much less likely to be related. A useful construct for thinking about such integration in the temporal domain is the "multisensory temporal binding window," a probabilistic function representing the likelihood that stimuli will be perceptually bound based on their relative timing. In this construct, multisensory stimuli presented simultaneously or at short delays are typically integrated into a unitary percept, whereas those presented at longer delays are less likely to be combined. Recent research from our laboratory examining this temporal window in children and adults suggests that it may mature quite late during the developmental process. These preliminary findings provided the foundation for the current experiment, which systematically documented the developmental progression of the multisensory temporal window by analyzing responses on a behavioral audiovisual simultaneity judgment task in subjects ranging from six to 23 years of age. The results reveal a decrease in window size with increasing age, and group analyses indicate that these changes extend well into adolescence (12-17 years). Together, these results illustrate a surprisingly protracted time course for the development of adult-like sensitivity to audiovisual asynchrony, and have important implications for the maturation of brain networks subserving the processing of basic (i.e., non-linguistic) auditory and visual stimuli, as well as for the design of idealized learning environments that allow for children to make maximal use of complementary sensory information from different modalities.

Keywords: development, maturation, auditory, visual, multisensory, temporal, simultaneity

Introduction

Combining information from multiple sensory modalities can dramatically influence human behavior and perception. The manifestations of multisensory integration have been investigated in adults using a vast array of behavioral and perceptual methodologies. For example, studies have shown that the pairing of stimuli from two or more sensory modalities can result in decreased simple reaction times (Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Hecht, Reiner, & Karni, 2008; Molholm, et al., 2002; Murray, Foxe, Higgins, Javitt, & Schroeder, 2001), shortened saccadic reaction times (Amlot, Walker, Driver, & Spence, 2003; Colonius & Arndt, 2001; Harrington & Peck, 1998), and an improved ability to detects targets (Frassinetti, Bolognini, & Ladavas, 2002; Lovelace, Stein, & Wallace, 2003). Complementary electrophysiological and neuroimaging studies have begun to identify the time course and neural bases of interactions and have revealed a complex and dynamic brain network underlying these interactions (Giard & Peronnet, 1999; Gondan, Niederhaus, Rosler, & Roder, 2005; Molholm, et al., 2002; Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007).

Over the last 30 years, research has shown that the multisensory interactions that generate these behavioral and perceptual benefits are critically dependent upon the relationships between the paired stimuli. For example, the so-called "temporal principle," first established in neurophysiological studies in animal models (Meredith & Stein, 1986), illustrated both the importance of stimulus timing in dictating the magnitude of the resultant interaction, and also that substantial interactions could be generated even when stimuli were separated by short intervals of time. Subsequent behavioral studies have reinforced this principle by showing that multisensory-mediated performance gains are robust at short temporal offsets and are reduced and ultimately eliminated when stimuli are significantly temporally misaligned (Corneil & Munoz, 1996; Frassinetti, et al., 2002). The extent of time over which multisensory interactions are highly likely to be produced has been referred to as the temporal window of multisensory integration or multisensory binding (Colonius & Diederich, 2004; Hairston, Burdette, Flowers, Wood, & Wallace, 2005; Hillock, Powers, & Wallace, submitted; Koppen & Spence, 2007; Navarra, et al., 2005; Noesselt, et al., 2007; Powers, et al., 2009; Spence & Squire, 2003; van Wassenhove, Grant, & Poeppel, 2007). It has been speculated that the purpose of such a window is that it enables multisensory interactions to be flexibly specified, accommodating for differences in travel time for signals emanating from a common source (i.e., differences in the speed of propagation of visual and auditory signals) as well as for discrepancies in processing and relay time across the different sensory systems.

While the influence of stimulus timing on multisensory integration has been extensively studied in adults, surprisingly little is known about how this factor influences multisensory interactions in developing populations, particularly in children and adolescents. Infant work has established differences in the detection of interstimulus asynchrony in two to eight month old babies when compared with adults, with thresholds being on the order of four to five times larger in infants (Lewkowicz, 1996b). Additionally, prior research from our laboratory in 10 and 11 year olds and adults has indicated that differences in the size of the multisensory temporal window extend beyond the first decade of life (Hillock, et al., submitted). Thus, while differences in multisensory temporal processing have been identified at several time points during development, the time course for maturation of the multisensory temporal window still remains unclear. Accordingly, the current study seeks to examine this question by testing performance on a simultaneity judgment task in subjects over a broad range of ages (i.e., 6 to 23 years). These findings, which illustrate a surprisingly late maturation for multisensory temporal processing, are critically important for enriching our understanding of the distinctions between how children, adolescents and adults perceive and interact with the world. Furthermore, these results offer the first empirical report of the sensitive period for maturation of multisensory temporal processing, knowledge likely to be of major importance in identifying changes in the brain networks subserving multisensory temporal perception, and that may be critically important for the design and implementation of improved learning environments.

Methods

Subjects and Screenings

Participants were recruited via institutionally approved advertising materials. All subjects and parents/guardians of minors were assented and/or consented prior to study participation in accordance with the regulations and an approved protocol of the Vanderbilt Institutional Review Board (IRB).

Sixty typically developing individuals between the ages 6 and 23 years participated in the study and 45 were included in final analyses. Participants with a known history of neurological disorders and those that failed to demonstrate normal hearing (pure tone thresholds less than 25dB HL at octave frequencies from 250 to 8,000 Hz), normal or corrected-to-normal vision

(Snellen, 20/20 - 2 or better for each eye), and average or above-average intelligence (i.e., Kaufmann Brief Intelligence, second edition – composite intelligence quotient [IQ]) were excluded (n=2) (Kaufman & Kaufman, 2004). An additional subject was disqualified for failing to complete all parts of the assessment as well as six subjects based on their responses on the simultaneity judgment task. An explanation of disqualifying criteria from the simultaneity measure can be found in the "Data Analysis and Temporal Window Derivation" section below.

In addition to hearing, vision and IQ screenings, sight word reading ability was assessed with the Test of Word Reading Efficiency (TOWRE) (Torgesen, Wagner, & Rashotte, 1999) and household socioeconomic status (SES) was calculated using the Hollingshead "Four Factor Index of Social Status" (Hollingshead, 1975). No subjects were excluded on the basis of reading ability or SES. Administration of screening measures required approximately 45 minutes.

Stimuli and Experimental Design

For the audiovisual simultaneity judgment measure (adapted from Fujisaki et al., 2004) participants were seated in a quiet, dimly lit room approximately 60 cm from a high refresh rate computer monitor (NEC Multisync F3992 [160 Hz refresh capacity] set to 100 Hz). A white crosshair fixation marker (0.75 cm x 0.75 cm) appeared in the center of a black background on the computer screen for the entirety of the experiment. Auditory (10 ms duration, 1800Hz tone burst, 89 peak dB SPL [A weighted]) and visual (10 ms duration, white ring flash, outer diameter = 12 cm, inner diameter = 7.0 cm, area = 113.1 cm^2) stimulus pairs were presented in a randomly interleaved fashion at the following visual-auditory stimulus onset asynchronies (SOAs): 0, +/- 50, +/- 100, +/- 150, +/- 200, +/- 250, +/- 300, +/- 400 ms and +/- 500 ms (Figure 3.1). Positive values represent visual leading events whereas negative numbers indicate that the auditory cue preceded the visual stimulus. The SOAs used were strategically chosen based on prior research and pilot work and represent those that most adequately capture changes in simultaneity report for subjects within the age range studied (Hillock, et al., submitted; Lewkowicz, 1996a).

Auditory stimuli were presented via Sennheiser HD 265 linear supra-aural earphones and intensity was verified with a sound level meter (Larson Davis LxT2, 375A02 microphone). Signal duration and interstimulus delays were externally verified with an oscilloscope (Hameg Instruments HM507) within an error tolerance of 6 ms, the temporal fidelity enabled by the ASIO low latency driver used with MATLAB. Stimulus presentation and data logging was controlled using MATLAB 7.7.0 R2008b software.

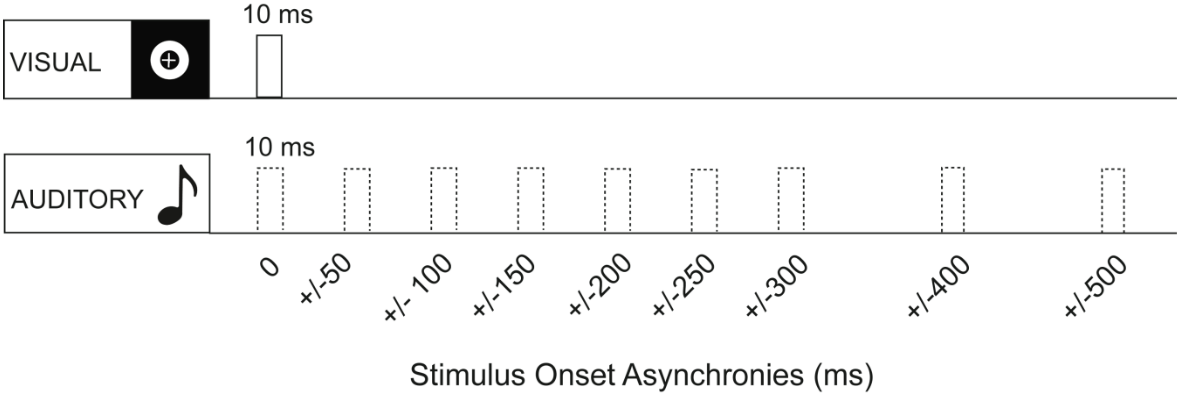


Figure 3.1. Simultaneity judgment task protocol. *Visual representation of the temporal structure between visual (ring flash) and auditory (tone pip) stimuli. The duration of each stimulus was 10 ms and the stimulus onset asynchrony for pairs ranged from 0 to 500 ms, where negative numbers indicate auditory leading visual presentations and positive numbers reflect auditory lagging conditions.*

For consistency, task instructions for all subjects were embedded in a story describing audiovisual communication in lighting bugs. This method facilitated understanding in younger children (Appendix 3.1). Behavioral judgments were recorded by pressing buttons with lightning bug images, blue (male) or red (female), which denoted simultaneous or successive auditory and visual stimulus presentations, respectively (Cedrus RB-530 response pad). Responses were counterbalanced across participants. Subjects were asked to respond as accurately as possible; speed was not emphasized in an effort to reduce errors. Prior to the assessment, a circumscribed set of pre-test questions was administered to verify understanding and a practice session (comprised of 5 trials) was completed. Subjects were given the option of repeating the practice up to two times before beginning the assessment. During the assessment, a total of 374 responses were collected (i.e., 22 samples at each SOA condition [17]). Trials were initiated one second after participants logged their response to the previous presentation. The assessment took approximately 10 - 15 minutes to complete. Participants were informed of their progress toward completion of the task via visual puzzles, which became progressively fuller. The puzzles appeared each time 25% of total trials were completed and participants were given the option of taking a short break (approximately 5 minutes) upon completion of 50% of trials (halfway through the experiment).

Temporal Window Derivation and Data Analysis

Temporal windows were derived in each individual from curves fitted to the mean probability of simultaneity report at each SOA. Two sigmoids were generated in MATLAB to the average simultaneity judgment values produced from responses at negative (-500 to 0 ms) and positive (0 to +500 ms) SOAs. The distributions were comprised of interpolated y values (probability of simultaneity report) at x values (SOA [ms]) ranging from -600 to + 600 ms in 0.1 ms increments. The temporal window was established as the width of each distribution (in ms) at ¾ maximum (probability of simultaneity report). Participants in which the probability of simultaneity report did not decrease to ¾ maximum (n=3) were excluded from analyses, as well as three outliers whose windows were more than two standard deviations from the group mean. A description of group composition appears below.

Data were divided into one of three groups based on subject age: children (range = 6-11 years, mean = 9.5 years), adolescents (range = 12-17 years, mean = 14.7 years), and adults (range = 18-23 years, mean = 21.2 years). Although 51 subjects met inclusion criteria, because more adolescents were recruited than children or adults, groups were restricted to 15 subjects each (n=45 total). Window size data from participants (randomly selected in groups with > 15 subjects) was compared using independent samples *t*-tests to obtain a global measure of group differences. A Pearson correlation was also performed between participant age and overall temporal window size to corroborate overall group analyses. For a more detailed depiction of contributions to group differences in window size, a multivariate repeated measures analysis of variance (rmANOVA) with a within-subjects factor of SOA condition (18 levels) and a between subjects factor of age group (3 levels) was performed. Greenhouse-Geisser corrections were used (where needed) to correct for dependence among the repeat measures within subjects. Planned comparisons were used to further illuminate the SOAs contributing to group differences, by examining differences in simultaneity report between children and adults and adolescents and adults at moderate and long SOAs. Independent samples *t*-tests were exclusively performed on SOAs of 150 ms and greater as prior findings from our lab indicated that differences in simultaneity report between children and adults were restricted to moderate and long SOAs (Hillock, et al., submitted).

To test for the possibility of rapid, within-session changes in performance (i.e., fatigue effects, learning effects), window sizes were computed in each individual on the first and last half of responses in the assessment. Because windows could not be derived in all subjects using the reduced number of trials, within session analyses were performed on groups comprised of 13 of the 15 subjects included in the tests described above. Paired samples *t*-tests were used for within-group comparisons of window sizes.

Results

Age-Related Differences in the Size of the Multisensory Temporal Binding Window

For the initial analyses, subjects were first divided into three large groups (n=16/ group) based on age (children [6–12 years], adolescents [13-17 years], adults [18-23 years]). Representative examples of the response distributions for the simultaneity judgment task are shown in figure 3.2 for one subject in each of these age groups. Differences in the width of the distribution are apparent, and when these distributions are used to create a singular measure of the multisensory temporal binding window (see methods), a marked difference between younger and older participants is revealed. Note that while each of these distributions is centered on 0 (i.e., simultaneity), the breadth of the distribution (and temporal window) for the adult participant is substantially narrower than for the adolescent or child participants. A comparison across groups (n=15 each) revealed a comparable pattern for the mean data, with no significant difference in mean window size between children and adolescents, but both groups differing significantly from adults. Thus, whereas average window size for children was 404 ms and for adolescents was 399 ms, this value was only 290 ms in the adults ($p< 0.01$ both tests (Figure 3.3). This group-based comparison was reinforced by correlation analyses, which showed a significant negative relationship between age and window size, $r = -0.403$, $p < 0.01$ (Figure 3.4). Data were best fitted with a polynomial line, suggesting that windows size is similar in children and young teens, and starts to shrink toward late adolescence into early adulthood.

Age-Related Difference in Simultaneity Report

In order to follow up and validate these group and individual comparisons that were performed using a global and singular measure of multisensory temporal processing (i.e., window size), additional analyses focused on examining the mean probability of simultaneity report across groups and SOAs. Significant main effects were seen for SOA condition (*F*(4.782, 200.865) = 153.138, *p* < 0.001), group (*F*(2, 42) = 6.017, *p* < 0.01), as well as a significant interaction between SOA and group, (*F*(9.565, 200.865) = 2.105, *p* < 0.05) (Figure 3.5). Further reinforcing the results of the group analyses focused on window size, these follow up measures indicated equivalent performance between children and adolescents, but significant differences between both children and adults and adolescents and adults. Tables 3.1 and 3.2 summarize these results. Overall, these findings show that children and adolescents are more likely to report moderately and long auditory leading stimulus pairs as simultaneous when compared with adults, suggesting substantial differences in the temporal constraints of multisensory binding at these ages.

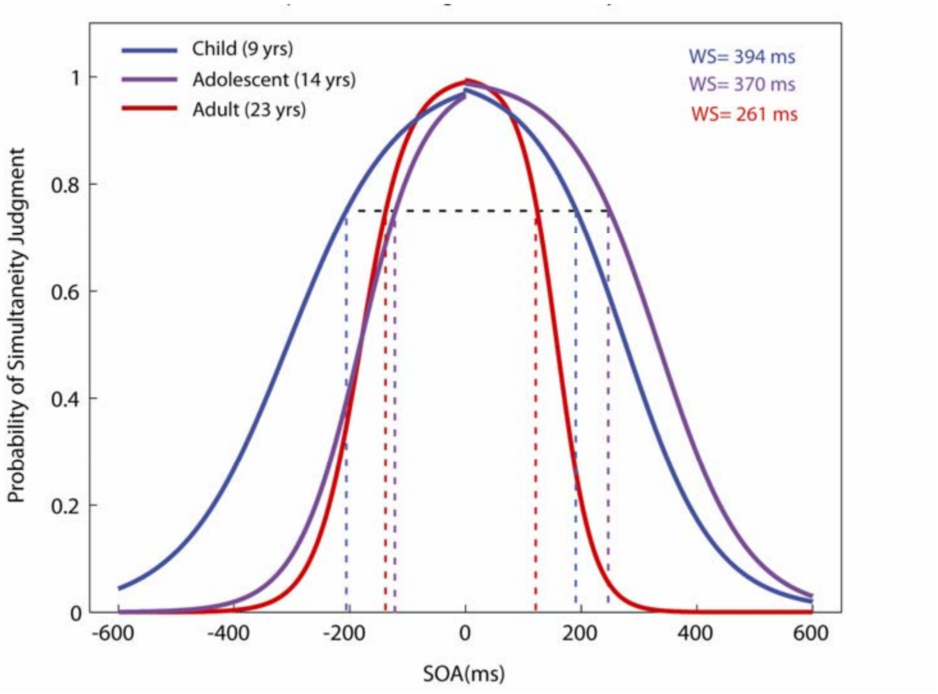


Figure 3.2. Individual subject data reveals that multisensory temporal binding windows are larger in children and adolescents than adults. *Younger subjects had winder windows than the adult, suggesting that they were less sensitive at detecting asynchrony. Horizontal dotted line indicates ¾ maximum (approximately .75), and vertical dotted lines delineate window boundaries for each individual.*

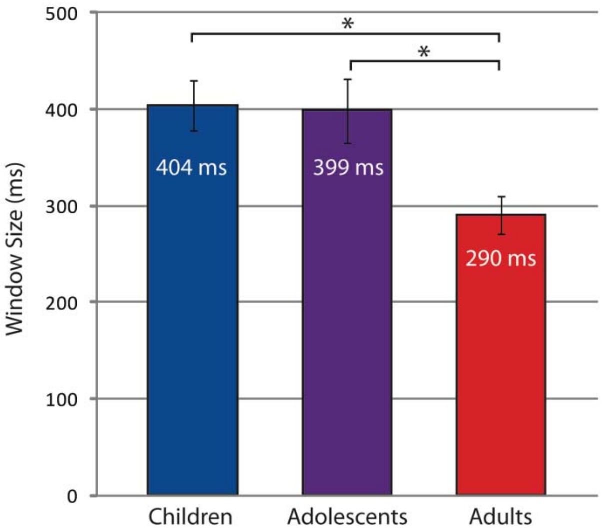


Figure 3.3. Mean window size is smaller in adults than in children and adolescents. *Bar graph displays mean window size for children (blue, left), adolescents (purple, middle) and adults (red, right)(n=15 subjects/ group). Children and adolescents have significantly smaller windows than adults, * = p< 0.01. Error bars indicate +/- one standard error of the mean (SEM).*

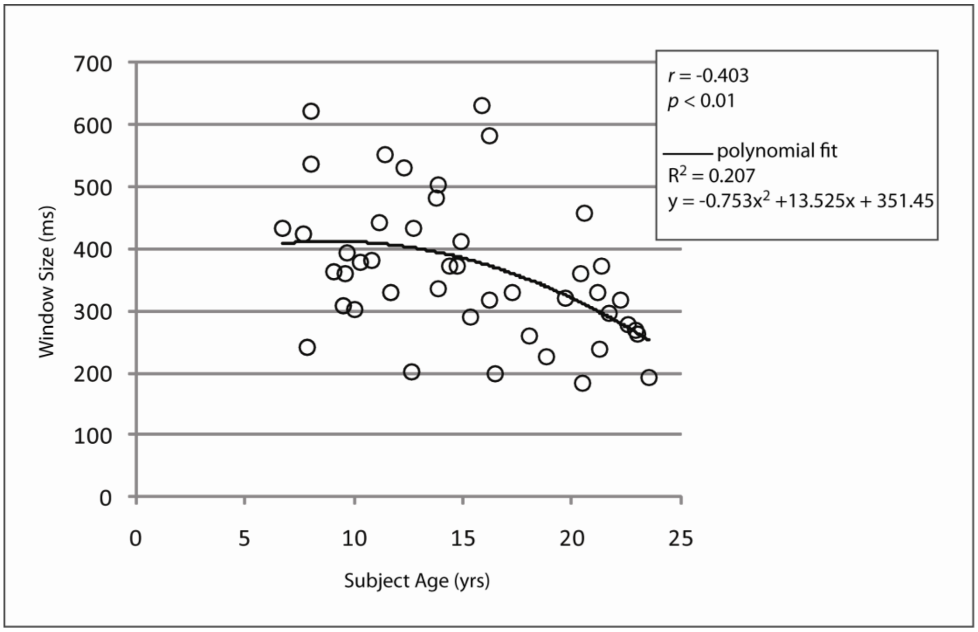


Figure 3.4. Multisensory temporal binding windows are smaller in older participants. *A significant negative correlation was observed indicating that older subjects have smaller windows. An exponential curve fit to window size data indicates that age accounts for 17.6% of variance in window size.*

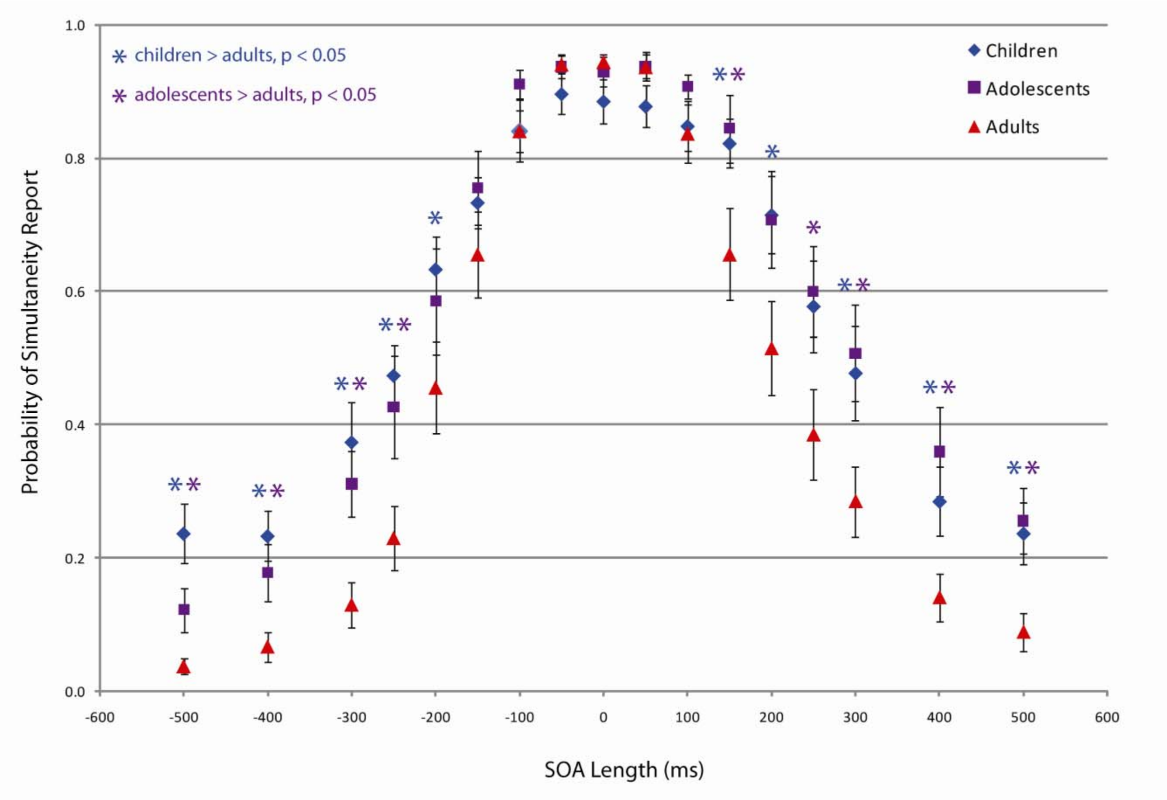


Figure 3.5. Children and adolescents are less sensitive to audiovisual asynchrony at moderate and long SOAs. *Graph shows mean probability of simultaneity report for each group (n=15 subjects/ group) at each SOA condition (-500 ms to + 500 ms). Planned comparisons revealed significant differences in probability of simultaneity report between children and adults and adolescents and adults at moderate and long SOAs, * = p < 0.05. Error bars = +/- one SEM.*

Table 3.1. Statistical table displaying differences in the probabilty of simultaneity report between children and adults. *Planned comparisons were performed at positive and negative SOAs from 150 to 500 ms. Analyses revealed that children are more likely to report stimuli as simultaneous at moderate and long SOAs (excluding- 150 and 250 ms).*

		Mean	SEM	t	df	p
-500	Children	0.237	0.044	4.405	15.986	*0.000
+	Adults	0.037	0.012			
-400	Children	0.233	0.037	3.831	23.007	*0.001
+	Adults	0.067	0.022			
-300	Children	0.374	0.059	3.585	22.656	*0.002
+	Adults	0.130	0.035			
-250	Children	0.474	0.045	3.697	28	*0.001
	Adults	0.230	0.049			
-200	Children	0.633	0.049	2.112	25.258	*0.045
+	Adults	0.456	0.069			
-150	Children	0.733	0.380	1.034	28	0.312
	Adults	0.656	0.065			
150	Children	0.822	0.037	2.135	28	*0.042
	Adults	0.656	0.069			
200	Children	0.715	0.059	2.184	28	*0.038
	Adults	0.515	0.070			
250	Children	0.578	0.069	1.982	28	0.057
	Adults	0.385	0.068			
300	Children	0.478	0.071	2.184	28	*0.038
	Adults	0.285	0.053			
400	Children	0.285	0.052	2.304	28	*0.029
	Adults	0.141	0.036			
500	Children	0.237	0.046	2.702	23.514	*0.013
+	Adults	0.089	0.029			

+ Equal variances not assumed.

Table 3.2. Comparisons of the probabilty of simultaneity judgment in adolescents and adults. *Planned comparisons at positive and negative SOAs from 150 to 500 ms indicated that the proportion of simultaneous reponses in higher in adolescents at positive and negative long SOAs and some moderate delays (-250, 150 ms).*

		Mean	SEM	t	df	p
-500	Adolescents	0.122	0.033	2.428	17.458	*0.026
+	Adults	0.037	0.012			
-400	Adolescents	0.178	0.042	2.315	21.293	*0.031
+	Adults	0.067	0.022			
-300	Adolescents	0.311	0.049	3.012	28	*0.005
	Adults	0.130	0.035			
-250	Adolescents	0.426	0.077	2.158	28	*0.04
	Adults	0.230	0.049			
-200	Adolescents	0.585	0.079	1.236	28	0.227
	Adults	0.456	0.069			
-150	Adolescents	0.756	0.055	1.174	28	0.250
	Adults	0.656	0.065			
150	Adolescents	0.844	0.051	2.206	28	*0.036
	Adults	0.656	0.069			
200	Adolescents	0.707	0.072	1.910	28	0.066
	Adults	0.515	0.070			
250	Adolescents	0.600	0.068	2.241	28	*0.033
	Adults	0.385	0.068			
300	Adolescents	0.507	0.072	2.486	28	*0.019
	Adults	0.285	0.053			
400	Adolescents	0.359	0.067	2.866	21.199	*0.009
+	Adults	0.141	0.036			
500	Adolescents	0.256	0.049	2.936	22.855	*0.007
+	Adults	0.089	0.029			

+ Equal variances not assumed.

Within-Session Changes

In an effort to determine whether there were within-session performance changes that might be indicative of non-stimulus dependent effects such as learning and fatigue, datasets were split into those acquired during the first and last half of trials during each session. No significant difference was seen between these two epochs, $p > 0.05$ all tests. Rather, performance appeared highly consistent throughout the assessment. This finding was further supported by the presence of strong correlations between within-session window sizes in children ($r = 0.74$), adolescents ($r = 0.78$) and adults ($r = .70$), $p < 0.01$ all tests. If fatigue was an issue, one might expect to see increased noise in data, which could result in relative widening of the window in the latter half of the experiment or even in difficulties in window derivation because of increased variability. Hence, the observed pattern of results support a sensory basis for group differences in simultaneity perception, arguing against the contribution of procedural learning, fatigue, or attention in producing observed effects.

Discussion

The current study represents the first empirical report of changes in multisensory temporal processing from early childhood through early adulthood, and provides compelling evidence that changes in perception of multisensory temporal relations extend well into adolescence. Results demonstrate that sensitivity to audiovisual temporal asynchrony increases with age, and that adults are less likely to bind more temporally disparate multisensory stimuli than younger participants. The study extends findings from our earlier work contrasting simultaneity judgment in children and adults, and (using standard methods of correction for multiple comparisons) results are entirely consistent with those of the prior study.

An important consideration when interpreting findings in developmental studies such as this is the potential influence of attention and vigilance on subject performance. Prior work comparing accuracy and reaction times on Continuous Performance Tests (CPT) reports that the capacity to sustain attention increases with age (Lin, Hsiao, & Chen, 1999; Rebok, et al., 1997), and is reduced in children and adolescents relative to adults (Cornblatt, Risch, Faris, Friedman, & Erlenmeyer-Kimling, 1988; McKay, Halperin, Schwartz, & Sharma, 1994). Thus, the period in development associated with decreases in the capacity to focus attention overlaps with the time period in which reductions in sensitivity to multisensory temporal asynchrony are observed. To mitigate the potential influence of attention on performance on the current task, several experimental design considerations were employed. First, the overall duration of the psychophysical assessment was short, lasting only 10-15 minutes, and was divided into two parts separated by a short break. Second, visual puzzles were used to maintain engagement for the duration of the task. Arguing against a strong attentionally-based interpretation of the results is the fact that post hoc comparisons of window size during the course of testing revealed little to no performance changes between early and late intervals.

Clearly the most surprising finding of the current study is the protracted achievement of the adult state. This raises the obvious question - why do audiovisual temporal processing capabilities mature so late? Perhaps experience with real world multisensory stimulus relations drives the development of a “mature” temporal window, and the requisite experience with these relations has yet to be fully realized. In a related manner, perhaps the integrative process is highly dependent on the nature of the stimuli being combined, and that the combination of arbitrary stimuli like those used in the current study are not fully reflective of the integration of more ethologically relevant stimuli (see discussion of speech stimuli below). Future studies can target this interpretation by using more realistic auditory and visual stimuli, such as a ball colliding with a substrate. Alternatively, one might posit a more brain-based developmental explanation for these results, and argue that the late maturation of the integrative process is a result of the delayed maturation of the brain networks responsible for the appropriate temporal calculations. Several of the more prominent areas likely to be involved in encoding multisensory stimulus timing include the insula (Bushara, Grafman, & Hallett, 2001), superior temporal cortex (Calvert, Hansen, Iversen, & Brammer, 2001; Macaluso, George, Dolan, Spence, & Driver, 2004; Noesselt, et al., 2007; Powers, Hevey, & Wallace, Unpublished results; Raij, Uutela, & Hari, 2000), and temporo-occipito-parietal junction (Raij, et al., 2000), although primary sensory areas have also been proposed to play an important role (Dhamala, Assisi, Jirsa, Steinberg, & Kelso, 2007; Noesselt, et al., 2007). As "associational" areas these regions are among the latest to mature in the cortical hierarchy. As an example, structural MRI studies have shown that regions of the superior temporal cortex show gray matter density and total volume changes up to 20 years of age (Gogtay, et al., 2004). Continued changes in system maturation and organization into early adulthood could provide a basis for age-related differences in multisensory temporal processing extending into the adolescent phase.

The implications of the current results are far-reaching and impact domains ranging from educational strategies through reading and speech and on to developmental disabilities. An extended temporal binding window for simple stimuli like those used in the current study raise interesting questions about how children perceive rapidly changing stimuli in their environment and in which the tendency will be for the increased binding of "inappropriate" audiovisual pairs. Should educational environments strive to slow the pace of stimulus delivery in an effort to promote appropriate matches? Or should efforts be made to narrow the temporal binding window with the assumption that such narrowing will improve perceptual (and cognitive) capacities? Recently our laboratory has shown that significant plasticity can be engendered in the size of the adult multisensory temporal binding window through classic perceptual learning approaches (Powers et al. 2009) – approaches that could be readily adapted for use in developing populations. The use of low-level (and arbitrary) stimuli in the current study raises the question of the how higher-order multisensory stimuli are bound in these same populations, and whether the temporal networks for multisensory processes such as speech and reading are comparably delayed in their developmental maturation. Perhaps the strong ethological imperatives for these cognitive processes (as well as their unique but overlapping neural bases) drive an accelerated timeline for their maturation. Parallel experiments are ongoing to examine the temporal constraints for audiovisual speech stimuli. Finally, emerging evidence suggests that multisensory temporal processing may be preferentially impacted in developmental disabilities such as autism and dyslexia (Foss-Feig, et al., 2010; Hairston, et al., 2005). Hence, a greater knowledge of the normative time course for multisensory temporal maturation provides an essential foundation within which to interpret these data, and perhaps more importantly may result in the creation of more sensitive diagnostic tools for the evaluation of these disabilities.

Acknowledgments

This work was supported by the Vanderbilt Kennedy Center for Research on Human Development and the American Speech Language and Hearing Association. We would like to thank Drs. Linda Hood, Wesley Grantham and Bruce McCandliss for their intellectual contributions and editorial assistance. We also wish to thank Matthew Hevey, Nicky Hackett, Olivia Broaddus and Haley Eberly for their assistance with data collection.

References

Amlot, R., Walker, R., Driver, J., & Spence, C. (2003). Multimodal visual-somatosensory integration in saccade generation. *Neuropsychologia, 41*(1), 1-15.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci, 21*(1), 300-304.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438.

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys, 63*(1), 126-147.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci, 16*(6), 1000-1009.

Cornblatt, B. A., Risch, N. J., Faris, G., Friedman, D., & Erlenmeyer-Kimling, L. (1988). The continuous performance test, identical pairs version (CPT-IP): I. new findings about sustained attention in normal families. *Psychiatry Research, 26*(2), 223-238.

Corneil, B. D., & Munoz, D. P. (1996). The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci, 16*(24), 8193-8207.

Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L., & Kelso, J. A. (2007). Multisensory integration for timing engages different brain networks. *Neuroimage, 34*(2), 764-773.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res, 143*(4), 480-487.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., et al. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res, 203*(2), 381-389.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res, 147*(3), 332-343.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Giard, M. H., & Peronnet, F. (1999). Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Journal of Cognitive Neuroscience, 11*(5), 473-490.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A, 101*(21), 8174-8179.

Gondan, M., Niederhaus, B., Rosler, F., & Roder, B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept Psychophys, 67*(4), 713-726.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res, 122*(2), 247-252.

Hecht, D., Reiner, M., & Karni, A. (2008). Multisensory enhancement: gains in choice and in simple response times. *Exp Brain Res, 189*(2), 133-143.

Hillock, A. R., Powers, A. R., 3rd, & Wallace, M. T. (submitted). Binding of sights and sounds: Age-related changes in multisensory temporal processing.

Hollingshead, A. B. (1975). Four factor index of social status. Yale University.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaugman Brief Intelligence Test, Second Edition*. Minneapolis, MN.

Koppen, C., & Spence, C. (2007). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Res, 1186*, 224-232.

Lewkowicz, D. J. (1996a). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*(5), 1094-1106.

Lewkowicz, D. J. (1996b). Perception of auditory-visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform, 22*(5), 1094-1106.

Lin, C. C., Hsiao, C. K., & Chen, W. J. (1999). Development of sustained attention assessed using the continuous performance test among children 6-15 years of age. *J Abnorm Child Psychol, 27*(5), 403-412.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res Cogn Brain Res, 17*(2), 447-453.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage, 21*(2), 725-732.

McKay, K. E., Halperin, J. M., Schwartz, S. T., & Sharma, V. (1994). Developmental analysis of three aspects of information processing: Sustained attention, selective attention, and response organization. *Developmental Neuropsychology, 10*(2), 121 - 132.

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

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res, 14*(1), 115-128.

Murray, M. M., Foxe, J. J., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia, 39*(8), 828-844.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res Cogn Brain Res, 25*(2), 499-507.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., et al. (2007). Audiovisual Temporal Correspondence Modulates Human Multisensory Superior Temporal Sulcus Plus Primary Sensory Cortices. *J. Neurosci., 27*(42), 11431-11441.

Powers, A. R., 3rd, Hevey, M., & Wallace, M. T. (Unpublished results). Neural correlates of multisensory perceptual training.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28*(2), 617-625.

Rebok, G. W., Smith, C. B., Pascualvaca, D. M., Mirsky, A. F., Anthony, B. J., & Kellam, S. G. (1997). Developmental changes in attentional performance in urban children from eight to thirteen years. *Child Neuropsychology, 3*(1), 28 - 46.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45*(3), 561-571.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Curr Biol, 13*(13), R519-521.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of word reading efficiency*. Austin, TX.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Percept Psychophys, 67*(3), 531-544.

Appendix

Appendix 3.1. Story script for simultaneity judgment task.

Title: The Great Bug Escape

Page 1: Earlier today at the Nashville Zoo the boy and girl lightning bugs got out of their cages and got mixed up. They need to be returned to their separate homes by tomorrow morning.

Page 2: It's getting late and has become dark outside. You can't see the bugs. The only way you can tell them apart is by their chirps and flashes. The lightning bugs' tails look like this circle…

Page 3: The boy lightning bugs chirp and flash their tail at exactly the **SAME** time.

Page 4: The girl lightning bugs chirp BEFORE or AFTER they flash their tails. The girl bugs NEVER chirp and flash at the same time.

Page 5: Your job is to help us sort the lightning bugs. When you hear the bug chirp and see a flash at the **SAME** time, press the button for the **boy** bug.

Page 6: When the chirp and flash **do NOT** happen at exactly the same time, press the button for the **girl** bug. Do your best! Thank you for your help!

<u>*Instructions*</u>: During this task look straight ahead at the screen. Keep your eyes focused on the plus sign. A white ring will appear and a tone will be played. Press the button to respond after both the ring and tone are played. Get ready. Press any button to start the game.

CHAPTER IV

A DEVELOPMENTAL STUDY OF THE TEMPORAL CONSTRAINTS FOR AUDIOVISUAL SPEECH BINDING

Abstract

In a typical communicative exchange, both visual and auditory cues *can* facilitate comprehension. The striking influence of vision on processing of an acoustic message has been well documented in the literature using speech in noise paradigms and perceptual illusions such as the McGurk effect. During a McGurk task, incongruent auditory and visual speech syllables are paired to induce the perception of a novel, fused token – a process reflective of the integration of these different sensory cues. Consistent with other multisensory interactions, the strength of the illusion is temporally constrained such that report of fused tokens decreases with increased temporal offset between auditory and visual stimuli. Whereas the McGurk effect and its temporal boundaries have been well characterized in adults, it has not been evaluated during early life, where previous research has shown there to be marked maturational changes in the temporal structure of how basic, non-speech multisensory stimuli are processed. Extending these findings into the speech realm is particularly important given the critical role of multisensory interactions in promoting the development of effective communication skills. To explore this question, perception of the McGurk illusion was measured in subjects ranging from 6 to 23 years of age using audiovisual pairings with temporal delays ranging from 0 to 500 ms (visual preceding auditory). In contrast to previous work using low-level audiovisual stimuli, although there was an age-dependent change in the frequency of perceptual fusion for the McGurk stimuli, no difference was seen in the temporal constraints for audiovisual binding for these speech stimuli across age. These results have important implications for the development

of overall behavioral and perceptual competencies, each of which are strongly influenced by multisensory interactions.

Keywords: development, maturation, auditory, visual, multisensory, temporal, McGurk effect

Introduction

Visual Influences on Speech Perception and the Concept of a Temporal Window

The influence of visual cues on speech comprehension has been well characterized in a number of behavioral studies. These studies have reinforced that the information conferred by a talker's face and lip movements can increase the intelligibility of the auditory signal (Macleod & Summerfield, 1987; Schwartz, Berthommier, & Savariaux, 2004; Sumby & Pollack, 1954), as well as improve speech detection thresholds in noise (Grant & Seitz, 2000). Whereas the visual contribution to speech recognition is nominal in quiet environments and at favorable signal-to-noise ratios (SNRs), it generally becomes more robust at higher noise levels and in reverberant listening environments when the auditory signal is degraded (Erber, 1969; Sumby & Pollack, 1954).

While the synergistic use of auditory and visual cues typically provides an adaptive benefit in the realm of speech comprehension, audiovisual interactions can also result in illusory percepts. Although such illusions have less ethological relevance, they can be used as important tools to index interactive processes taking place within speech networks. One of the most convincing and informative of these illusions is the McGurk effect (McGurk & Macdonald, 1976). In this paradigm, the presentation of incongruent visual (glottal - /ga/, /ka/) and auditory (bilabial - /ba/, /pa/) speech tokens often results in report of an intermediary fused (velor) target such as

/ta/ or /da/, reflecting an active integration of the visual and auditory elements. As with normal speech processes, the strength of the McGurk effect is highly dependent upon the salience of the auditory cues, and reliance on visual information increases when the auditory signal is ambiguous and/or degraded (Fixmer & Hawkins, 1998; Sekiyama, Kanno, Miura, & Sugita, 2003; Sekiyama & Tohkura, 1991). In addition to being modulated by stimulus efficacy, the McGurk effect is also influenced by the temporal relationship between the auditory and visual cues, such that increasing asynchrony decreases the strength of the illusion (Jones & Jarick, 2006; Massaro, Cohen, & Smeele, 1996; Soto-Faraco & Alsius, 2009; van Wassenhove, Grant, & Poeppel, 2007). Nonetheless, there is a window of time over which individuals experience a high degree of fusion, which typically spans asynchronies ranging from -50 to +200 ms, where negative and positive values indicate auditory leading and lagging conditions, respectively (Jones & Callan, 2003; Massaro, et al., 1996; Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove, et al., 2007). The concept of a temporal window for multisensory integration is not unique to speech stimuli, and appears to extend not only to other forms of audiovisual stimuli (i.e., simple flashes and beeps), but also to reflect interactive processes ranging from the level of the single neuron through a variety of behaviors and perceptions (Frens, Van Opstal, & Van der Willigen, 1995; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hillock, Powers, & Wallace, submitted; Meredith, Nemitz, & Stein, 1987; Powers, Hillock, & Wallace, 2009; Stein, Huneycutt, & Meredith, 1988; Vroomen, Keetels, de Gelder, & Bertelson, 2004; Zampini, Guest, Shore, & Spence, 2005).

Changes in Audiovisual Integration during Development

While there is now a wealth of data on how stimulus characteristics shape the multisensory integrative process in adults, much less is known about how these interactions appear and mature during early life. A number of questions come immediately to the forefront. Do infants and children employ similar strategies to adults in establishing what stimuli belong together? How (and when) do we become sensitive to the statistical relationships between multisensory stimulus complexes? Finally, how do these relationships change as the individual sensory systems develop (each at different rates)? Prior work examining temporal processing in infants has established that by four months of age babies can detect tempo, rhythm and synchrony in the auditory and visual modalities and can use these cues for crossmodal matching (Bahrick, 1983; Bahrick, 1987, 1988; Dodd, 1979; Lewkowicz, 1986, 1992, 1996, 2000; Mendelson & Ferland, 1982; Spelke, 1979). Furthermore, evidence suggests that 4.5 to 5 month old infants are not only able to relate between the auditory and visual modalities, but also have the capacity to actively integrate sensory cues from different modalities, as demonstrated by the appearance of responses indicative of perception of the McGurk illusion (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997). Thus, research suggests that infants are remarkably sensitive to stimulus characteristics such as temporal synchrony, can match multisensory stimuli based on amodal attributes (e.g., intensity, location, duration, etc.), and demonstrate the capacity to integrate multisensory stimuli (Bahrick & Pickens, 1994). However, despite these remarkable abilities, multisensory development is far from complete in the infant stage.

Studies focusing on later development suggest that multisensory processing abilities are further refined during early childhood. For example, it has been shown that pre-school (i.e., 3-4 year old) and school-aged (i.e., 4 ½ to 10 year old) children are less likely to experience the McGurk illusion than more mature subjects (i.e., 18-40 years of age) (Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; McGurk & Macdonald, 1976). This finding has been extended in a recent study that showed fewer McGurk illusions in 5-9 year old children when compared with 10-14 and 15-19-year olds (Tremblay, et al., 2007). Hence, there is evidence of marked changes in the nature of audiovisual interactions to simultaneously presented speech stimuli during early childhood, but the details of how the temporal structure of these stimuli influences responses remain to be elucidated.

Developmental Changes in Sensitivity to Interstimulus Temporal Asynchrony

While our knowledge concerning the developmental antecedents leading up to mature multisensory speech processes has grown appreciably in recent years, we still know relatively little about the maturation of the temporal factors likely to play an integral role in these processes (given their well established importance in adults). What work that has been done in the temporal domain has focused on more reduced (i.e., non-speech) stimuli, and has given us important insights into the developmental building blocks likely to result in mature speech representations. For example, comparison of behavioral report in adults and gaze maintenance findings in babies revealed differences in detection of asynchrony between a video track of a ball bouncing on a platform and the corresponding sound of the collision (Lewkowicz, 1996). Whereas the mean adult threshold for detection of asynchrony in visual leading presentations was slightly more than 100 ms, infants appear to have thresholds that are more than four times those of adults. Extending this to older ages, recent work from our laboratory indicates that these changes in multisensory temporal processing extend beyond the first decade of life. Comparison between children and adults on an audiovisual simultaneity judgment task revealed significant differences in the probability of reporting simultaneity at moderate and long asynchronies; 10 and 11 year olds were more likely to bind these stimuli (i.e., report simultaneity) compared with adults (Hillock, et al., submitted). While these results suggest that changes in the temporal architecture of audiovisual speech integration may extend into childhood, thresholds for the detection of asynchrony in speech and non-speech stimuli are known to differ considerably in adults (Dixon &

Spitz, 1980). Adults are reportedly more tolerant to temporal mismatch between audiovisual speech stimuli when compared with non-speech cues (i.e., a hammer hitting a platform). Given these differences, it remains plausible that the timeline for the maturation of audiovisual speech and non-speech processes, as well as their underlying neural correlates, may differ. The current study provides the first systematic attempt to examine the temporal architecture of audiovisual binding processes for speech stimuli across a wide window of development, and provides a follow up to a complementary study designed to examine these same questions for non-speech stimuli.

Methods

Subjects and Screenings

Participants were recruited via institution approved advertising materials and all participants and parents/ guardians of minors were assented and consented prior to study participation in accordance with the regulations of the Vanderbilt Institutional Review Board (IRB).

Sixty typically developing individuals between the ages 6 and 23 years participated and 44 were included in analyses. Thirteen participants were excluded based on their responses on McGurk trials in the audiovisual speech task. A description of disqualifying criteria appears in the “Data Analysis” section below. Participants with known neurological issues, hearing or uncorrected vision loss, below average intellectual ability and those that failed to complete all parts of the study were also disqualified (n=3). The measures and criterion adopted to determine study eligibility included the following: hearing evaluation (pure tone thresholds < 25dB HL at octave frequencies from 250 to 8,000 Hz), visual acuity (Snellen, 20/20 - 2 or better for each eye), and intellectual ability (i.e., Kaufmann Brief Intelligence, second edition – average or above average composite IQ [standard score >= 85]) (Kaufman & Kaufman, 2004).

Additionally, sight word reading ability was assessed using the Test of Word Reading Efficiency (TOWRE) (Torgesen, Wagner, & Rashotte, 1999) and household socioeconomic status (SES) was calculated using the Hollingshead "Four Factor Index of Social Status" (Hollingshead, 1975). Subjects were not disqualified on the basis of SES or reading ability. Screening measures took approximately 45 minutes to complete.

Stimuli and Experimental Design

For the computer based speech tasks, participants were seated in a quiet, dimly lit room approximately 60 cm from a high refresh rate computer monitor (NEC Multisync F3992, refresh rate set to 100 Hz). To degrade the auditory cues thereby increasing the frequency of McGurk illusion perception (Soto-Faraco & Alsius, 2009; Sumby & Pollack, 1954), auditory tokens were embedded in white noise. Because speech perception studies have reported increased interference of auditory noise on speech discrimination in young children relative to older children and adults (Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Fallon, Trehub, & Schneider, 2000; Hall, Grose, Buss, & Dev, 2002; Johnson, 2000; Talarico, et al., 2007), an adaptive Levitt staircase was run prior to the audiovisual speech task to equate auditory discrimination abilities across subjects (Levitt, 1971). A three down, one up procedure was employed to determine the amount of noise needed to produce unisensory auditory accuracy scores of 79% correct. The four syllables used in the staircase and audiovisual speech task to follow were natural productions from a native English-speaking male (i.e., /ba/, /ga/, /da/, /tha/, mean duration = 511 ms, range = 463 – 555 ms). The tokens used included those paired to induce the McGurk illusion, /ba/ and /ga/, as well as the most commonly reported (fused) McGurk percepts, /da/ and /tha/.

Before beginning the staircase, subjects were asked to point to the button depicting each consonant-vowel (CV) (recited aloud by the experimenter) to verify their ability to correctly associate the visual and auditory representations of the stimuli. After this pre-test, the staircase was initiated and participants responded to each trial by pressing one of four buttons depicting the CVs heard (Cedrus RB-530 response pad). The level of the speech stimuli was fixed (56 dB SPL +/- 0.5 dB SPL, all tokens) and the intensity of the broadband noise was varied. The starting level of the noise was 50 dB SPL and the initial step size (i.e., amount by which noise level is changed) of 4 dB was decreased to 2 dB after the second reversal. Two independent tracks were run simultaneously (trials interleaved and alternating) and the final noise level was given by the average of the thresholds produced by each track. However, in instances where only one track terminated normally (8 reversals were not completed, n=2), the noise level was set based on the threshold produced from the terminated track only. Total test time for the staircase was approximately 5 minutes. Auditory stimuli were presented via Sennheiser (HD 265 linear) supra-aural earphones and stimulus intensity was verified with a sound level meter (Larson Davis LxT2, 375A02 microphone). Stimulus presentation and data logging for both the staircase and audiovisual speech task was controlled using Matrix Laboratory (MATLAB) 7.7.0 R2008b software.

The audiovisual speech task was administered immediately after completion of the staircase and assessed discrimination of unisensory and multisensory (congruent and incongruent) CV syllables. Trials included in this assessment were as follows: 1) auditory only (9 samples/ token [4 tokens - /ba/, /ga/, /da/, /tha/]), 2) visual only (9 samples/ token [4 tokens]), 3) multisensory congruent (9 samples/ token [4 tokens], and 4) multisensory incongruent (18 samples/ SOA [1 token – auditory /ba/, visual /ga/]; 9 SOAs). Instructions for the audiovisual task were embedded in a computer-generated story to promote understanding in younger participants

(Appendix 4.1). After the story, all participants completed a practice comprised of seven trials. Subjects could repeat the practice up to two times before beginning the task. The speech tokens and response options were identical to those employed in the staircase. A white crosshair fixation marker (1.9 cm x 1.9 cm) appeared in the center of a black background on the computer screen prior to and following each trial. Visual only and audiovisual tracks showed the male talker mouthing the speech syllables (dimensions: forehead to chin (mouth closed) = 9.5 cm, forehead to chin (mouth open) = 10.2 cm). Unisensory auditory stimuli were presented with a static image of the talker's face. All congruent multisensory pairs were presented simultaneously. McGurk pairs (auditory /ba/, visual /ga/) were presented at the following visual leading stimulus onset asynchronies (SOAs): 0, 50, 100, 150, 200, 250, 300, 400 ms and 500 ms (Figure 4.1). Stimulus duration and interstimulus delays were externally verified with an oscilloscope within an error tolerance of 5 ms (Hameg Instruments HM507).

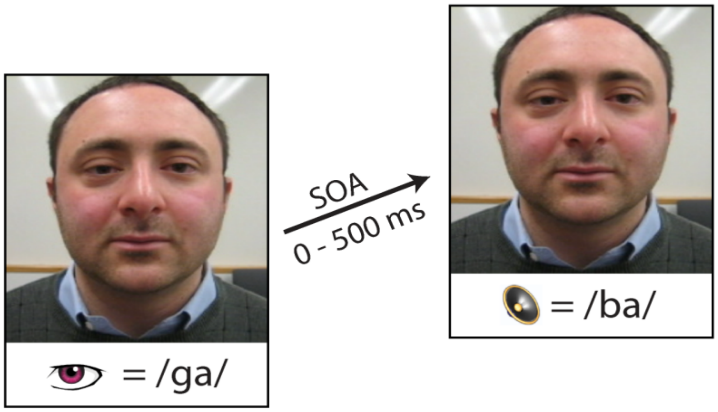


Figure 4.1. Visual representation of the trial structure for incongruent McGurk stimuli. *For McGurk trials, the auditory track was presented either simultaneously with the onset of lip movement, or after at the following SOAs: 50, 100, 150, 200, 250, 300, 400, 500 ms.*

Trials were randomly interleaved and broadband noise was played during all presentations (unisensory and multisensory) at the pre-determined level individually established via the staircase procedure. Unisensory auditory and visual trials and synchronous, congruent audiovisual pairs were each sampled 9 times, whereas incongruent tokens were sampled 18 times at each SOA. Trials were initiated approximately 1 second after participants logged their response to the preceding presentation. A total of 270 trials were presented during the experiment, which lasted approximately 30-40 minutes long. Visual puzzles appeared four times throughout the experiment; they became progressively fuller and displayed the percent of trials completed (i.e., 25%, 50%, 75% and 100%). Participants were given the option of a short break (i.e., approximately 5 minutes) at each interval.

Data Analysis

Performance on the incongruent multisensory (McGurk) trials was evaluated to determine if subjects met the criterion for inclusion in the analyses. The probability of fusion (report of /da/ or /tha/) was calculated for collapsed SOAs and individuals that reported fusions on 33% of trials or less (n=9), chance levels (1- /ba/, 2- /ga/, 3- /da/ *or* /tha/), were excluded. Additionally, mean fusion was assessed at each individual SOA and subjects that failed to exhibit temporal dependency of multisensory interactions, reduced fusion at moderate to long SOAs (200-500ms) relative to short-moderate delays (0-200 ms), were also eliminated.

Once inclusion measures were complete, the connection between age and unisensory (i.e., visual only, auditory only) and congruent multisensory control trials was assessed via Pearson correlations (n=44). These were used to identify the potential influence of audibility, lipreading proficiency and task understanding on performance on incongruent McGurk trials. Next, differences in performance on McGurk pairs were assessed via a bivariate correlation between age and the probability of fusion (collapsed SOAs), and a partial correlation that controlled for the influence of visual only ability.

To further define differences in integration, responses on McGurk trials were compared across groups characterized by subject age (i.e., children [6-11 years], adolescents [12-17 years] and adults [18-23 years]). To equalize the number of participants in each group (n=14), data from two randomly selected adolescents were excluded from group measures. Independent samples t-tests were used to compare mean fusion across groups for collapsed SOAs. To establish the specific asynchronies producing group differences, a multivariate repeated measures ANOVA (within subjects' factor of SOA condition [9 levels] and between subjects factor of age group) and post hoc tests were performed. Follow-up independent samples *t*-tests were used to determine the specific SOAs producing group differences. Criterion values on *t*-tests were adjusted using the Holm correction for multiple comparisons. Unfused responses on McGurk trials were descriptively compared for collapsed SOAs to evaluate potential differences in the rate of auditory or visual report between groups.

To establish a global metric of temporal processing for each age group, the multisensory temporal binding window was computed from group-averaged responses and compared using descriptive statistics. To derive windows, the mean probability of fusion was calculated at each SOA and responses were normalized to the 0 ms SOA condition (separately for each group). Thus, the value at the 0 ms SOA was multiplied by a real number which increased the proportion of fused trials to 1; SOAs from 50 to 500 ms were all multiplied by that constant. Numbers that exceeded 1 (max = 1.097) were rounded down to 1.0 to allow fitting of the sigmoid function. The sigmoid was generated in MATLAB to the normalized fusion values from 0 to + 500 ms. Distributions were comprised of interpolated y values (probability of fusion) at x values (SOA [ms]) from 0 to + 600 ms in 0.1 ms increments. The temporal window was established as the width of each distribution at ¾ maximum.

Lastly, within session changes were assessed to determine the potential influence of fatigue or learning effects in groups. Accordingly, the probability of fusion (averaged SOAs) was calculated for each subject for the first and last half of trials. The mean rate of fusion on the early and later parts of the assessment were compared in each group using paired samples *t*-tests.

Results

Performance on Control Conditions Reveals an Age-Related Change in Lip-reading Ability

In order to properly interpret the developmental changes in McGurk perception as a function age, it is first necessary to examine changes in unisensory (i.e., visual alone, auditory alone) performance associated with the task, as well as multisensory controls (e.g., congruent visual and auditory tokens) that can index developmental changes outside of the realm of multisensory integration. Since efforts were made to control for stimulus audibility across subjects and ages by use of the staircasing procedure, it was of little surprise that there was no correlation between age and auditory alone performance ($p > 0.05$). SNRs were properly adjusted to equalize performance (mean = - 6.2 dB SNR; range = + 8.5 to – 12.2 dB SNR). In addition, congruent multisensory performance accuracy was universally high, with no correlation being present between age and performance ($p > 0.05$). Equivalent performance on this control suggests that even the youngest subjects understood the task. In contrast to auditory only and bimodal congruent trials, a significant relationship was observed between lipreading ability and age, $F(2, 39) = 3.428$, $p < 0.05$ (Figure 4.2), with performance increasing as a function of age.

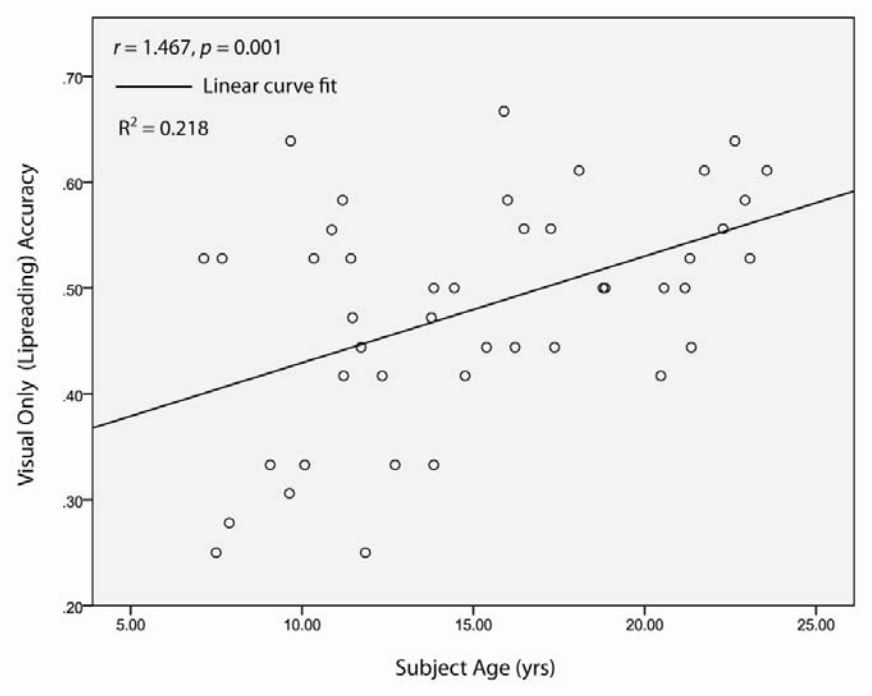


Figure 4.2. Lipreading ability improves with age.

Performance on McGurk Trials Reveals an Age-Dependent Effect on Illusion Perception

When a correlational analysis was performed between overall reports of fusion (i.e., perception of the McGurk illusion on incongruent trials) and age, a significant positive relationship was revealed (Figure 4.3A). Hence, older individuals show a higher likelihood of reporting the illusion on these trials. Given the age-dependent changes in visual-only (i.e., lip reading) performance established above, partial correlation analyses were done in order to control for the influence of lipreading ability. Results revealed a moderately strong and significant positive relationship between probability of fusion and subject age (*pr* = 0.385, *p* < 0.05), indicating that illusion perception is increased in older subjects even after accounting for differences in visual performance. While the rate of fusion was comparable for children and adolescents, children fused less than adults (*t*(26) = -3.582, p = 0.001) as did adolescents (*t*(26) =

-2.316, $p < 0.05$) (Figure 4.3B). Overall, adults perceived the illusion on approximately 20% more trials than children, with adolescents showing intermediate levels of fusion.

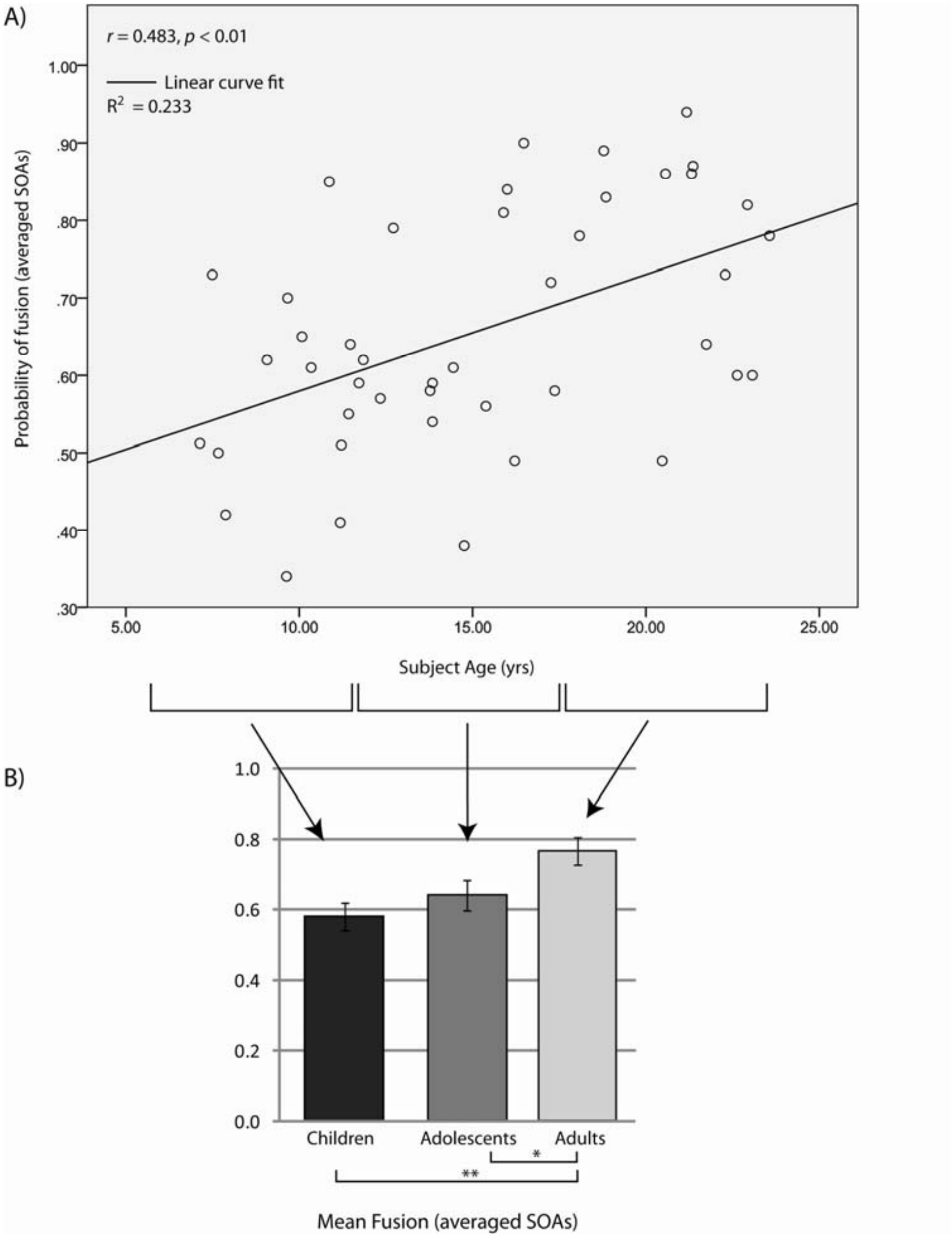


Figure 4.3. Older subjects experience the McGurk illusion more frequently than younger participants. A) *A bivariate correlation between age and illusion perception revealed a significant positive relationship indicating that older subjects are more likely to perceive the McGurk effect (n= 44).* ***B)*** *Bar graphs show significant decreases in fusion in children and adolescents relative to adults (n = 14/ group). Error bars indicate +/- one standard error of the mean (SEM). * $p < 0.05$; ** $p <= 0.001$*

The Temporal Constraints for Illusion Perception Differs Little Across Age

Whereas these initial analyses focused on overall reports of fusion collapsed across all of the tested SOAs, subsequent analyses sought to examine changes in the temporal structure of the multisensory integration indexed by the McGurk illusion. In order to do this, data from incongruent McGurk trials was then compared across groups (n=14 each) at individual SOAs for each possible response option (i.e., fusion, visual token, auditory token). For fusion responses, these analyses revealed significant main effects of both group ($F(2, 39) = 6.267$, $p < 0.01$) and SOA ($F(5.323, 207.591) = 54.790$, $p < 0.001$) after Geisser-Greenhouse correction, but no significant interaction ($p < 0.05$). Follow up analyses indicated that these effects were due to decreased fusion in children relative to adults at short and moderate SOAs (i.e., those from 0 to 300 ms, excluding 250 ms: $p < 0.05$ all tests; Holm correction for multiple comparisons) (Figure 4.4A/B). In contrast, adolescents did not differ from either group ($p > 0.05$). These results serve to reinforce the results of the collapsed analyses described above, in supporting a decreased incidence of fusion report for children at short to intermediate SOAs. Surprisingly, the lack of an interaction effect suggests that the temporal constraints of the fusion process do not change significantly among the groups tested (see additional analyses below).

In addition to measuring fusion, analyses also focused on characterizing responses when subjects failed to report the illusory percept. To characterize these unfused responses, the proportion of reports of the auditory and visual tokens during incongruent trials were compared across groups for collapsed SOAs. Children and adolescents were over three times more likely to report the auditory token (> 30%) than the visual token (< 9%) on these trials. In contrast, in adults responses were almost balanced across modalities (auditory = 13%, visual = 10%) (Figure 4.5). This raises an obvious question as to whether lipreading ability may influence report type. As expected, a correlation analysis revealed a strong relationship between visual only performance and auditory responses (*r* = -0.530, *p* = 0.000) such that poorer lipreaders (younger groups) are more likely to report the auditory token on unfused McGurk trials.

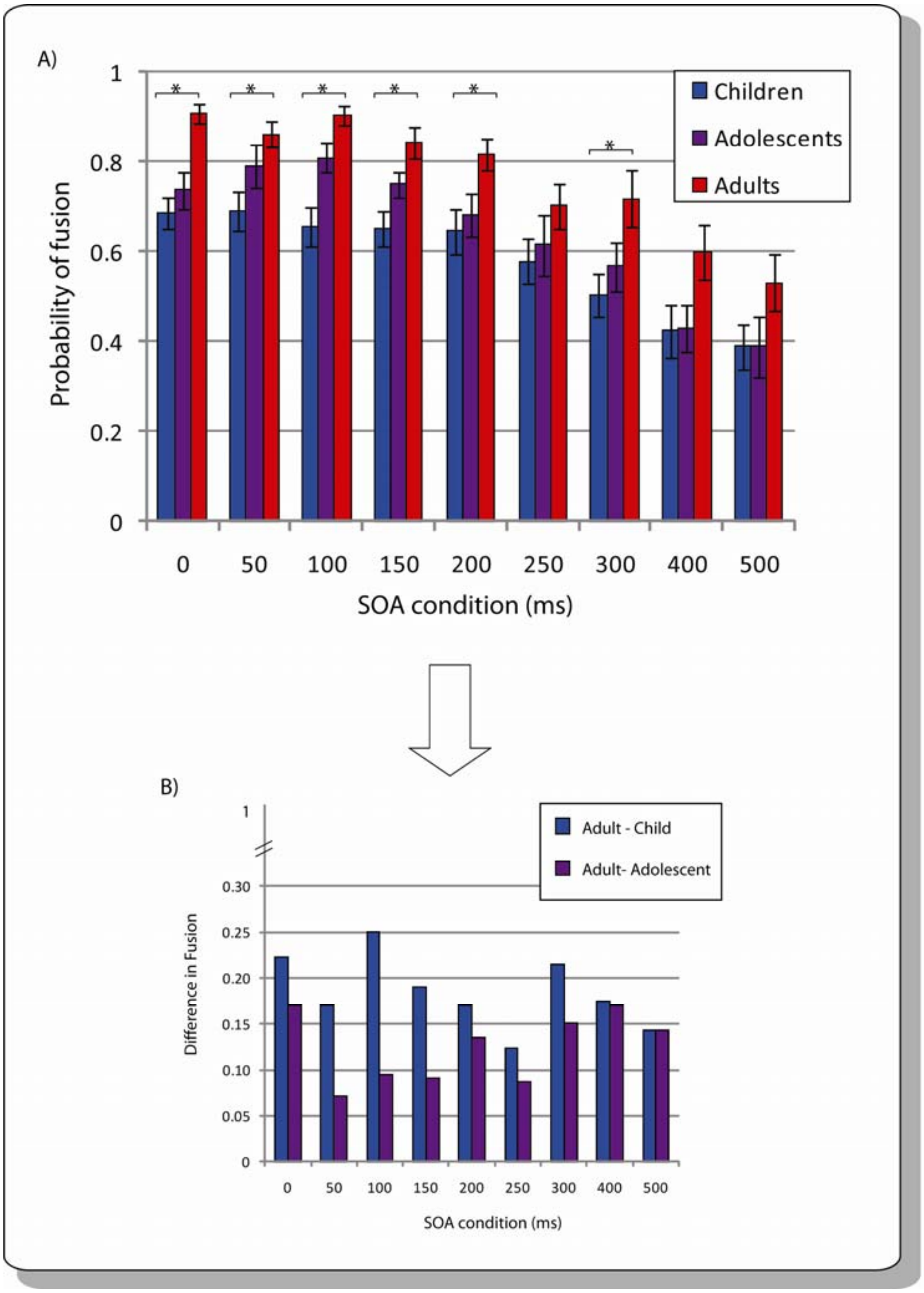


Figure 4.4. Children fuse incongruent audiovisual stimuli less frequently than adults at short and moderate SOAs. A) *Bar graph indicates mean fusion in each group (children= blue; adolescents = purple; and adults = red). Results indicated significant group differences between children and adults at SOAs from 0 – 300 ms (excluding 250 ms). Error bars represent +/- one SEM. * = p < 0.05.* ***B)*** *Graphs displaying group difference scores (Adults- Children= blue; Adults- Adolescents = purple).*

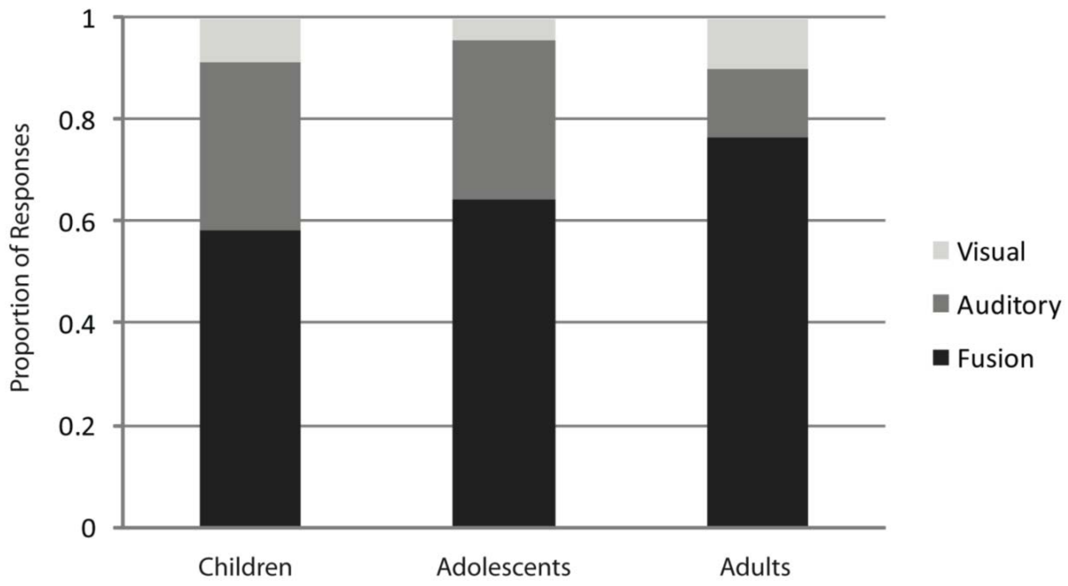


Figure 4.5. Children and adolescents are more likely than adults to report the auditory token on unfused McGurk trials. *Whereas children and adolescents reported the auditory token more than three times more often than the visual cue on unfused trials, this difference was reduced in adults.*

While these analyses show age-related changes in multisensory binding as a function of age, they fail to specifically address the question as to how the temporal binding window for speech stimuli changes with development. To assay potential age-related differences in temporal window size, response distributions were created for each of the groups showing fusion responses as a function of SOA (Figure 4.6). As is evident, these distributions are strikingly similar. When they are distilled down to a singular measure of the size of the multisensory temporal binding window (i.e., width at 75% maximum, see methods), this lack of difference is further reinforced. Hence, mean window sizes were nearly identical across groups (children = 352.6 ms, adolescents = 353.8 ms and adults = 353.1 ms), indicating that children, adolescents and adults are equally sensitive to audiovisual asynchrony for speech.

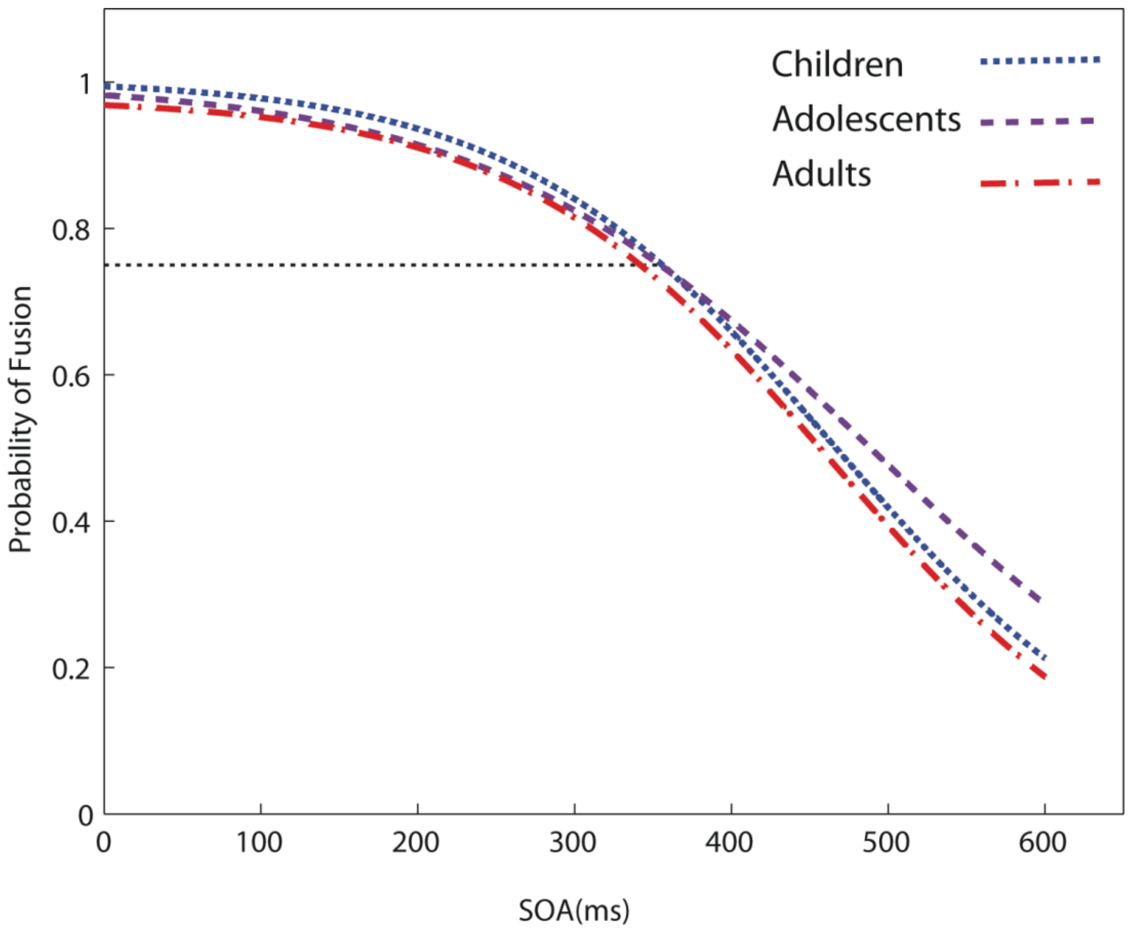


Figure 4.6. Audiovisual temporal binding windows for speech are similar in children, adolescents and adults. *Window size was comparable in all three groups, suggesting that temporal processes governing multisensory integration are mature even in young children. Curves have been vertically shifted for clearer presentation.*

Within-Session Analyses Reveal Little Difference in Fatigue Across Groups

In an effort to assay whether differences in performance were due to fatigue or global changes in attentional processes, comparisons were made between fusion responses obtained during the first and last half of the assessment. Results of paired samples *t*-tests were not significant for any of the groups ($p > 0.05$ all tests), suggesting that performance (as measured via rate of illusion perception) was consistent throughout the experiment, and arguing against any age-dependent differences in fatigue or attention.

Discussion

The results of the current study show two main effects. First, and as has been previously reported (Massaro, 1984; Massaro, et al., 1986; McGurk & Macdonald, 1976; Tremblay, et al., 2007), perceptual fusions characteristic of the McGurk effect become increasingly common at later ages, and this effect appears to be partly attributable to the increasing influence of visual stimuli on speech processing as a function of development. Second, and quite surprising given findings from our prior work using non-linguistic audiovisual stimuli (Chapters I and II), the temporal structure for the binding of the auditory and visual elements of the speech stimuli did not change appreciably during the maturational process. These results suggest either that the contraction of the binding window for speech-related stimuli is rapid and occurs very early in development (i.e., prior to age 6), or that the temporal boundaries for the integration of communication signals are fixed from birth. The potential basis for group differences in reports of fusion, the lack of age effects on temporal window size, and potential generalizability of these findings is discussed.

Despite substantial experience in all subjects with speech, significant age-related differences in the frequency of fusion were noted. Correlational analyses revealed that at least a portion of this effect is directly attributable to differences in visual performance (i.e., lip reading skill), as indexed via responses on visual only trials. Such a result complements reports from recent modeling studies, which suggest that multisensory integration arises from the combination of statistically weighted unisensory cues, and in which weights are inversely proportional to stimulus reliability (Ernst, 2007; Helbig & Ernst, 2007; Witten & Knudsen, 2005). Generalizing this to the current work, it can be assumed that the visual signal is less reliable (and hence less heavily weighted) for children, increasing their relative weighting of the auditory signal. The developmental transition from auditory to visual modality dominance may also contribute to decreased visual weighting in younger subjects. Prior work has reported increased attention and preferential processing of auditory signals in infants and young children relative to adults, which is reflected in the decreased influence of visual stimuli on perception (Robinson & Sloutsky, 2004; Sloutsky & Robinson, 2008). For example, when provided with visual and auditory information as to the location of a target, responses of four-year-olds were generally consistent with the auditory cue whereas adults' responses were more reflective of the visual cue (Robinson & Sloutsky, 2004). In the current study, increased processing of the auditory signal by younger participants may have further inflated the relative weight of the auditory cue subsequently decreasing reports of fusion or the visual token on McGurk trials. Future research should examine developmental shifts in cue weighting by systematically manipulating auditory and visual cue reliability. Such studies have been performed in other modalities, where optimization of cue weighting reportedly emerges in middle childhood (8-10 years) (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Jones, Bedford, & Braddick, 2008).

While fusion rate was clearly influenced by age in the current study (as in prior work – see Massaro, 1984; Massaro et al., 1986; McGurk & McDonald, 1976; Tremblay et al., 2007), there was little evidence here for a change in the temporal bounds within which the integrative process is taking place. As alluded to earlier, this could reflect either a narrowing of this binding window prior to the ages studied here, or a fixed window size for speech-related signals. Regardless, either possibility is likely to be the result of the enormous imperative placed on communication and the prolific increase in speech perception and production in early development. By 10-12 months of age infants have become markedly tuned to the phonetic contrasts within the native language, and it is postulated that this canalization process enables more sophisticated processing of acoustic and phonetic features of the native language (Werker & Yeung, 2005). Hence, it is entirely plausible that the temporal binding window for speech stimuli is consolidated around this time, when babies become sensitized to the array of sounds upon which language is derived. In support of this concept is the marked similarities in the duration of English phonemes (i.e. 30 – 300 ms) (Crystal & House, 1981) and the size of the multisensory temporal binding window (i.e., approximately 350 ms). Thus, the temporal structure of these elemental communicative building blocks may dictate the limits of the audiovisual binding, at least for speech-related stimuli.

Indeed, these results for speech-related stimuli are a bit surprising given earlier findings from our lab showing that the binding window for low level audiovisual stimuli (i.e., flashes and tone pips) matures slowly, with changes extending into the adolescent period (Chapters I and II). Assuming that the temporal window for speech stimuli is plastic, this more rapid window consolidation may be related to the system specialization for language, a process that may be dictated by an accelerated maturation of the neural structures subserving speech processes. In an event-related potentials (ERP) study involving 3-16 year olds, Pang and Taylor (2000) reported that the N1c ERP component to speech was consistent with the adult morphology by 7-8 years of age, but that this same component was not clearly identifiable until 11-12 years for tones, after which time its amplitude continued to increase into adulthood. This finding may be due to developmental differences in the neural generators for these different stimuli, which may be biologically tuned for speech because of its importance for communication.

Based on the current findings, it is postulated that distinct processes lead to the attainment of mature multisensory integration abilities (as indexed through the frequency of reports of the McGurk effect) and multisensory temporal processing (as indexed through window size measures). Although the temporal window was mature even in our youngest group, children were less susceptible to the illusion when compared with adults. This latter finding may be reflective of inverse effectiveness, a property of multisensory interactions whereby the magnitude of integration (e.g., fusion) is dictated by the strength of unisensory cues and weaker stimuli are more likely to produce integration (Meredith & Stein, 1986). Auditory dominance effects in younger subjects could lead to increased saliency of the auditory speech signal and subsequently decreased fusion. Such findings have implications for overall functioning given the role of audiovisual interactions in promoting stimulus detection, speech recognition, etc. Maturational differences may be particularly apparent in noisy settings such as the classroom, where reductions in audiovisual integration render children and adolescents less capable of managing the deleterious effects of masking noise on speech understanding. On the other hand, multisensory temporal processing appears fully developed by age 6, which may be critically important in establishing the foundation for the development of reading skills. Abnormal multisensory temporal processing has been reported in individuals with reading impairment (Hairston, Burdette, Flowers, Wood, & Wallace, 2005; Rose, Feldman, Jankowski, & Futterweit, 1999), and it has been speculated that reductions in multisensory temporal resolution may lead to inappropriate correspondences between graphemes and phonemes. The current study suggests that the basis for the development of normal reading skills are in place at a very early age, perhaps even before formal instruction begins. However, future work should further explore the link between multisensory temporal processing and reading ability, and audiovisual integration and speech recognition. Multisensory based training has been shown to produce lasting changes in audiovisual temporal processing in adults (Powers, et al., 2009), and perhaps such methods may also prove useful in strengthening or accelerating reading ability, or remediating central auditory processing disorders that can cause discrimination deficits and increased noise sensitivity.

Acknowledgments

This work was supported by the Vanderbilt Kennedy Center for Research on Human Development and the American Speech Language and Hearing Association. We would like to thank Drs. Linda Hood, Wesley Grantham and Bruce McCandliss as well as Kathryn Guillot for their intellectual contributions and editorial assistance. We also wish to thank Matthew Hevey, Nicky Hackett, Olivia Broaddus and Haley Eberly for their assistance with data collection.

References

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development, 6*, 429-451.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior & Development, 10*, 387-416.

Bahrick, L. E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev, 59*(1), 197-209.

Bahrick, L. E., & Pickens, J. N. (1994). Amodal Relations: The basis for intermodal perception and learning in infancy. In D. L. Lewkowicz & R. Lickliter (Eds.), *The Development of Intersensory Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220.

Crystal, T. H., & House, A. S. (1981). Segmental durations in connected speech signals. *The Journal of the Acoustical Society of America, 69*(S1), S82-S83.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*(4), 478-484.

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.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J Speech Hear Res, 12*(2), 423-425.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J Vis, 7*(5), 7 1-14.

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.

Fixmer, E., & Hawkins, S. (1998). *The influence of quality of information on the McGurk effect.* Paper presented at the Australian Workshop on Auditory-Visual Speech Processing.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys, 57*(6), 802-816.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr Biol, 18*(9), 694-698.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am, 108*(3 Pt 1), 1197-1208.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hall, J. W. I., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear and Hearing, 23*(2), 159-165.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Exp Brain Res, 179*(4), 595-606.

Hillock, A. R., Powers, A. R., 3rd, & Wallace, M. T. (submitted). Binding of sights and sounds: Age-related changes in multisensory temporal processing.

Hollingshead, A. B. (1975). Four factor index of social status. Yale University.

Johnson, C. E. (2000). Children's Phoneme Identification in Reverberation and Noise. *Journal of Speech, Language & Hearing Research, 43*(1), 144.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *NeuroReport, 14*(8), 1129-1133.

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Exp Brain Res, 174*(3), 588-594.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaugman Brief Intelligence Test, Second Edition*. Minneapolis, MN.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2), 467-477.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development, 9*(3), 335-353.

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*(3), 297-324.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*(5), 1094-1106.

Lewkowicz, D. J. (2000). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development, 71*(5), 1241-1257.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131 - 141.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev, 55*(5), 1777-1788.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am, 100*(3), 1777-1786.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology, 41*(1), 93-113.

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Mendelson, M. J., & Ferland, M. B. (1982). Auditory-Visual Transfer in Four-Month-Old Infants. *Child Development, 53*(4), 1022-1027.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci, 7*(10), 3215-3229.

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

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol, 18*(9), 689-693.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Dev, 75*(5), 1387-1401.

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Futterweit, L. R. (1999). Visual and auditory temporal processing, cross-modal transfer, and reading. *J Learn Disabil, 32*(3), 256-266.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Percept Psychophys, 59*(3), 347-357.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition, 93*(2), B69-78.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res, 47*(3), 277-287.

Sekiyama, K., & Tohkura, Y. i. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America, 90*(4), 1797-1805.

Sloutsky, V. M., & Robinson, C. W. (2008). The Role of Words and Sounds in Infants' Visual Processing: From Overshadowing to Attentional Tuning. *Cognitive Science, 32*(2), 342-365.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *J Exp Psychol Hum Percept Perform, 35*(2), 580-587.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626-636.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res, 448*(2), 355-358.

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.

Talarico, M., Abdilla, G., Aliferis, M., Balazic, I., Giaprakis, I., Stefanakis, T., et al. (2007). Effect of age and cognition on childhood speech in noise perception abilities. *Audiol Neurootol, 12*(1), 13-19.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of word reading efficiency*. Austin, TX.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PLoS One, 2*(1), e742.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences, 9*(11), 519-527.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron, 48*(3), 489-496.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Percept Psychophys, 67*(3), 531-544.

Appendix

Appendix 4.1. Story script for audiovisual speech task.

Title: Talking Made Sense

Page 1: Billy goes to the movies every Friday. He usually arrives 10 minutes early to get a good seat so that he can see and hear everything.

Page 2: Last Friday Billy arrived a little late and almost all of the seats were taken. The seat he found was in the very back of the theater and it was hard for him to understand from where he was sitting.

Page 3: Please help Billy figure out what speech sounds the man in this movie is saying.

Instructions: You will see a plus sign to remind you to pay attention just before each movie clip starts. Watch the man's lips moving and listen to the syllables he makes. During this task, sometimes you will hear him talking but his lips won't move. Sometimes you will see his lips moving but not hear his voice. Other times you will hear him and see his lips moving. After each movie clip is finished, press the button that shows what he said. If you aren't sure, take a guess. Do you best! Press any button to start the game.

CHAPTER V

GENERAL DISCUSSION

Summary of Results

The overall goal of this dissertation was to delineate the timeline for the maturation of multisensory temporal processing for both speech and nonspeech stimuli. Findings from the series of studies reported have revealed that the age at which adult-like multisensory temporal processing is realized differs based on stimulus type. In Chapter II, it was established that changes in the temporal window for binding basic (i.e., non-speech) audiovisual stimuli persist beyond the first decade of life, with windows being significantly enlarged in 10-11 year old children. The subsequent study described in Chapter III extended these findings by detailing the developmental trajectory for window narrowing from early childhood to early adulthood in subjects 6 to 23 years of age. Comparison of window size across age groups indicated enlarged windows in the two youngest groups (6-12 years and 13-17 years) relative to adults, suggesting that adult-like temporal processing is not realized until late in the adolescent period. Much to our surprise, these findings sharply contrasted those generated in the same group of subjects for speech stimuli using a classic McGurk task (Chapter IV). Results of that study indicated that temporal windows to speech were mature even in our youngest group, as evidenced by nearly identical group averaged window sizes. Interestingly, however, further analysis of group performance revealed that were substantial decreases in the number of fusions reported by children relative to adults despite the similarity in temporal window size. Whereas children and adolescents bound basic auditory and visual signals to the same extent as adults within the temporal window, reports of integration were reduced in children compared to adults for speech stimuli. Thus, our findings indicate that the contraction of the temporal window for speech precedes that for basic non-speech stimuli, but that speech-related multisensory binding still matures over a fairly protracted time course. It is speculated that the accelerated time course to mature multisensory temporal processing of speech may be due to the system specialization for language and increased experience with speech versus nonspeech constructs. Overall, findings suggest potential differences in the neural circuitry mediating multisensory processing of speech and nonspeech stimuli and the time frame within which enhancements in perception of audiovisual speech and other environmental stimuli are fully attained.

In the sections to follow, some of the interpretational caveats of the work are addressed, and participants' performance on the Simultaneity Judgment and McGurk experiments is directly compared. Also, an experience-driven model of window contraction is proposed which address differences in the time frame for window consolidation across speech and nonspeech stimuli. Lastly, future study ideas are discussed and the implications of the work for clinical populations and the classroom environment are reviewed.

Key Study Considerations

Is Asymmetric Window Narrowing Consistently Observed for Basic Stimuli?

Results of the first Simultaneity Judgment study (Chapter II) showed an asymmetric window narrowing effect such that group differences in audiovisual simultaneity report were restricted to moderate and long negative SOAs. Findings suggested that 10 and 11 year-olds were more likely to report auditory leading audiovisual pairings as simultaneous than adults at greater temporal delays. Such an effect is also apparent in group-averaged distributions (although reduced) for positive SOAs; however, this difference was not significant after application of a statistical correction for multiple comparisons. A correction was applied because of the exploratory nature of post-hoc testing; we had no prior assumptions as to what SOAs might reveal group differences. On the contrary, in the follow-up study a restricted number of uncorrected planned comparisons were performed at SOAs strategically selected (-500 to -150ms, 150 to 500 ms) based on findings in the earlier experiment. As a result, significant group differences between both children and adolescents and adults were identified at moderate and long positive and negative SOAs. However, as was previously observed, differences were most robust for negative SOAs (and were larger in children than adolescents for the left side of the distribution). Hence, application of a statistical correction restricted significant differences to children and adults at negative SOAs from -250 to -500 ms, a finding that is entirely consistent with our earlier report of more rapid right-sided window contraction. Findings highlight the consistency of this effect and reinforce the potential role of experience in driving asymmetric maturational changes.

Can McGurk Fusions Accurately Define the Multisensory Temporal Binding Window?

One of the goals of the current work was to compare the maturational timeline for contraction of the audiovisual temporal binding window for nonspeech stimuli to that for speech stimuli. While prior work has indicated that audiovisual identification and simultaneity judgment tasks yield comparable results and presumably engage similar mechanisms, recently this interpretation has been challenged. Soto-Faraco and Alsius (2009) used a dual-task paradigm whereby participants reported both the synchrony relation between auditory and visual stimuli and the perceived identity of the utterance on each trial. A five alternative-forced-choice (AFC) response method (i.e., “bda”, “da”, “ba”, “dba”, and “other”) was used for identification and a synchrony/ asynchrony determination was made for the temporal task. van Wassenhove and colleagues (2007) used an alternate approach in which tasks were run consecutively; the identification experiment was completed prior to the simultaneity judgment task so that participants were ignorant to synchrony manipulations. Although the response format for the temporal task was similar to that described above, a three alternative-forced-choose (AFC) method was employed with the identification measure (i.e., “ba”, “ga”, “da”, tha”).

Using the dual-task five AFC measure Soto-Faraco and Alsius showed a large discrepancy in the size the temporal window for integration and asynchrony detection (Figure 5.1). Based on this they claim that participants can report illusions despite sensing temporal asynchrony and question the relationship of such measures. The conclusions of van Wassenhove and colleagues sharply contrast those of Soto-Faraco and Alsius, indicating that the temporal window for audiovisual integration is very similar to that for asynchrony detection (Figure 5.2A/B). Interactions are reportedly maintained across a span of approximately 200 ms. Study differences could be attributed to a number of factors (e.g., task order, response format). A more detailed view into the data reveals that the distribution presented by Soto-Faraco and Alsius was coarsely sampled relative to that of van Wassenhove and colleagues. Measurements were made at 13 SOAs from -640 to + 720 with adjacent steps sizes differing variably in the most extreme case by 240 ms. Thus, window boundaries are based on extrapolated data fit to widely spread points along the temporal continuum. Contrastingly, van Wassenhove and colleagues sampled responses at 29 SOAs from -467 to +467 ms enabling finer resolution of temporal changes, less extrapolation and presumably more accurate estimates of window size.

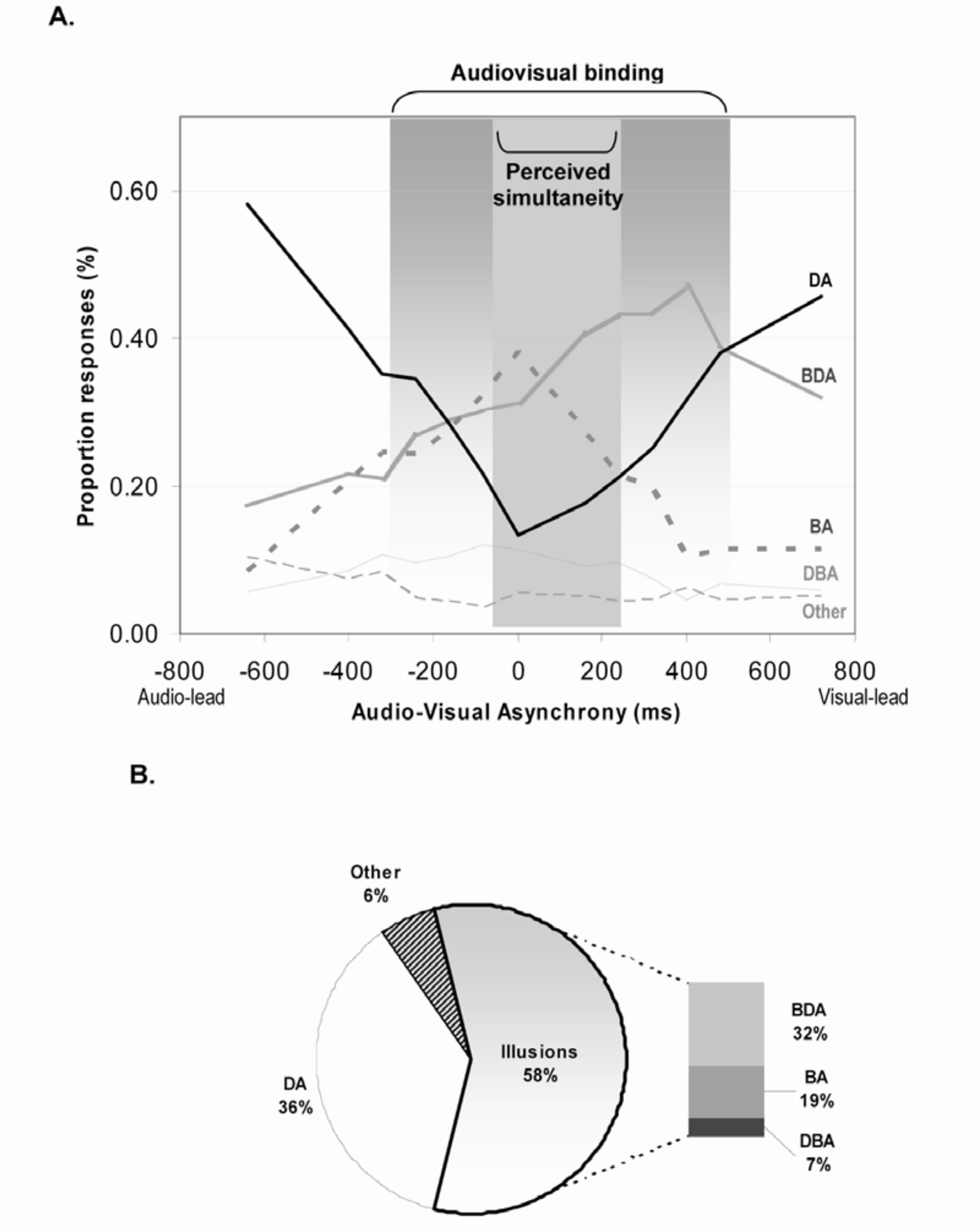


Figure 5.1. A case against the link between asynchrony detection and audiovisual integration. *Graph extracted from Soto-Faraco & Alsius, 2007.* ***A)*** *Depiction of the discrepancy in window sizes generated on a dual-task McGurk experiment characterizing asynchrony detection and speech identification responses. Authors report evidence of integration even in conditions where asynchrony is readily detected and suggest that these two measures reflect distinct multisensory processes.* ***B)*** *Pie chart showing responses characterized as fusions used to derive the audiovisual binding window.*

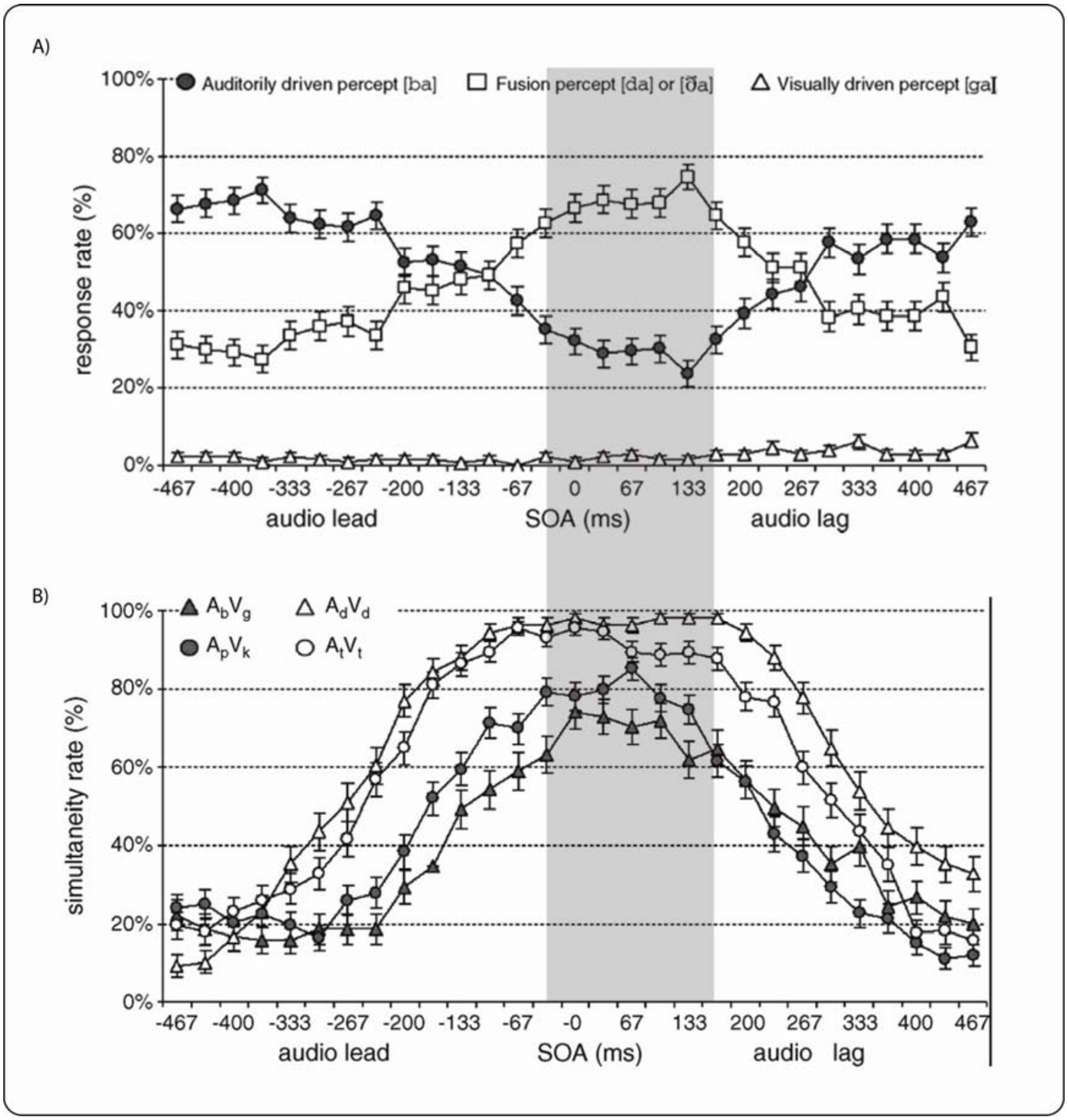


Figure 5.2. Evidence that multisensory temporal binding windows are similar for McGurk asynchrony detection and identification tasks. *Figures adapted from van Wassenhove, Grant & Poeppel, 2007.* ***A)*** *Temporal windows (indicated by the gray shaded region) derived from responses on the identification task and* ***B)*** *simultaneity judgment assessment are fairly equal (~200 ms each). Windows represent the span of time over which responses do not significantly differ from those obtained in the 0ms SOA condition.*

Differences may also arise from discrepancies in response options and in the categorization of integrated or fused responses. Classically, multisensory integration has been described as the binding of sensory information to produce a unitary percept. Thus, in the case of the McGurk effect, the purest form of multisensory integration is reflected by novel, fused percepts such as “da” or “tha”. In such cases, the system has negotiated discrepancies in the place of articulation of audio (bilabial) and visual (glottal) stimuli by inducing perception of an intermediary (alveolar) token. The approach of van Wassenhove was consistent with this interpretation, and only fused (i.e., “da”, “tha”) tokens were incorporated in estimates of the rate of illusion perception. In contrast, Soto-Faraco incorporated additional responses (i.e., “bda”, “dba”, “ba”) when calculating average illusory report (Figure 5.1B). In the strictest sense, such responses are not reflective of traditional fusion as they demonstrate the ability of participants to resolve unisensory components in audiovisual pairings. Inclusion of such tokens in estimates of illusion perception could result in an overestimation of the temporal window of audiovisual binding for speech relative to the synchrony window, and explain contradictions across studies.

While the method used by van Wassenhove and colleagues enables a clearer distinction between fused and unfused responses and a seemingly more accurate estimate of the temporal binding window, it is acknowledged that such an approach also has limitations given that it forces percepts into a restricted number of categories. These categories may not be entirely consistent with a participant’s sensory experience, but may instead represent the token that their percept most closely resembles. Some studies have avoided this by incorporating an “other” category (Munhall, Gribble, Sacco & Ward, 1996; Munhall & Tohkura, 1998), or by using an open-set format (McGurk & MacDonald, 1976; Sams, Manninen, Surakka, Helin & Kättö, 1998; Tremblay et al., 2007). However, such methods are also problematic given that the former does not allow participants to specify the nature of the percept, and the latter requires the experimenter to subjectively categorize participants’ responses.

The approach used in the study presented in Chapter IV is consistent with that of van Wassenhove and others (Sekiyama & Burnham, 2008). This method was chosen not only because it enables more objective response categorization, but because it reduces the number of response alternatives, making the task more manageable for younger participants. Additionally, it appears to be better correlated with asynchrony detection tasks, a finding of key importance for the current thesis, which aims not only to report the maturational timeline for the development of mature audiovisual temporal binding of speech and nonspeech stimuli, but to compare differences in the trajectory to realization of adult-like performance across measures. Based on our assessment of the studies above, we feel confident that the comparison across tasks is valid and that both audiovisual identification and simultaneity judgment measures engage common mechanisms of multisensory processing.

Does the McGurk Effect Accurately Capture the Temporal Window for Audiovisual "Speech" Processing?

Although McGurk studies utilize speech stimuli, the extent to which findings can be generalized to typical (congruent) multisensory syllables and to global speech processing has been questioned. Some evidence to suggest that audiovisual integration might differ for congruent and incongruent tokens comes from studies comparing audiovisual temporal asynchrony detection with matched and unmatched speech stimuli. Figure 5.2B illustrates the effect of congruity on multisensory temporal processing, indicating that individuals are more tolerant to matched tokens than unmatched pairs. Results suggest that estimations as to the size of the temporal binding window for speech using McGurk stimuli may slightly underestimate its breadth.

Despite behavioral indications that individuals are more adept at identifying asynchrony for incongruent speech, the neuroimaging literature suggests that the networks recruited for processing McGurk stimuli, natural congruent syllables and words are highly similar. These audiovisual speech stimuli all activate a host of cortical areas including the superior temporal gyrus/ sulcus, inferior frontal gyrus, extrastriate, premotor, parietal cortex, and primary auditory and visual cortices (Beauchamp, Nath & Pasalar; Callan et al., 2003; Calvert, Campbell & Brammer, 2000; Campbell, 2008; Jones & Callan, 2003; Stevenson & James, 2009). To our knowledge, only one structure amidst this network has been identified as differentially active to congruent versus incongruent audiovisual speech- the superior temporal sulcus (Calvert et al., 2000). To dissociate such an effect, Calvert and colleagues (2000) compared activations on conditions wherein the story mouthed in a video matched or differed from an audio track. Results indicated that comparable multisensory interaction effects were observed for both congruent and incongruent stimuli in all areas except STS, which showed enhanced activity to congruent pairings and decreased activity to incongruent combinations. Findings are further supported with evidence from unisensory auditory studies showing increased activation in anterior STS to intelligible speech relative to unintelligible or complex speech-like stimuli (Liebenthal, Binder, Spitzer, Possing & Medler, 2005; Scott, Blank, Rosen & Wise, 2000).

Nevertheless, there is extensive overlap in the structures activated to incongruent and congruent syllables and activations during syllable coding are highly analogous to those during word processing {Tan, 2000 #743; Rimol, 2006 #747}. Overall, findings suggest that the McGurk effect is a useful tool for indexing maturational changes in temporal structure of audiovisual speech binding.

Temporal Windows are Larger for Speech than Nonspeech Stimuli: A Comparison across Studies

The uniqueness of speech is evidenced when contrasting multisensory integration for different forms of audiovisual stimuli. In a seminal study by Dixon & Spitz (1980), asynchrony detection thresholds of adults were reportedly higher for speech-related audio and video tracks (i.e., person reading text) as compared to non-speech pairings (i.e., hammer hitting a platform), but also see

Conrey & Pisoni (2006). While multiple factors may contribute to these differences in temporal window size (e.g., task differences, signal duration), we attempted to investigate if windows generated to speech and nonspeech stimuli differed in our adult cohort and whether the assertion of winder windows for speech could be generalized to the developing population. Accordingly, group-averaged responses of children, adolescents and adults who satisfactorily completed both the Simultaneity Judgment and McGurk tasks were compared. (*Note: Group composition is overlapping but not entirely consistent with those established in Chapters IV and V*.) A common method of window derivation was used to analyze data from the two tasks wherein responses were averaged for positive SOAs (0 to 500ms), normalized to the veridically simultaneous condition, and windows were identified as the point along the abscissa where report of simultaneity or fusion perception decreased to three quarters maximum. *(See the data analysis section in Chapter IV for a fuller description of the window derivation method employed here.)* Results indicated that window sizes were larger for speech for all groups (children: speech = 357 ms, nonspeech= 224 ms; adolescents: speech = 353 ms, nonspeech = 249 ms; adults: speech = 320 ms, nonspeech= 156 ms) (Figure 5.3). Most importantly and as reinforced throughout the thesis, this discrepancy between asynchrony detection abilities for speech and nonspeech stimuli is less pronounced in children and adolescents, where windows for speech are likely to be mature while those for nonspeech are still consolidating. The delayed maturation of windows for nonspeech stimuli relative to speech may due to greater experience and system tuning for language, although another possibility is that speech requires less overall window contraction. In other words, if at 6 years of age windows are roughly equal in size (spanning approximately 350 ms), immaturities would be exclusively evident in the nonspeech realm, where further contraction is needed to reach the adult-like state given that windows are smaller for basic stimuli relative to speech.

These discrepancies in window sizes across stimuli may be related to the stronger semantic relationship for multimodal speech cues relative to nonspeech. Whereas ring flashes and tone pips are rather arbitrary pairings, the link between visual and auditory language is extremely meaningful and can promote integration even in early postnatal life (Burnham & Dodd, 2004; Rosenblum, Schmuckler & Johnson, 1997). Perhaps the brain is more apt to bind semantically congruent stimuli such as speech, subsequently creating greater tolerance for intersensory temporal asynchrony. Another intriguing theory is that the size of the binding window for speech is directed by the temporal structure of language. The average duration of the smallest speech units (i.e., phonemes) (Arai & Greenberg, 1997) is approximately 30-300 ms. These elements dictate the coarse temporal structure of language and the upper bound of their duration is remarkably similar to the size of the temporal binding window for speech, as has been insinuated by others (van Wassenhove, Grant & Poeppel, 2007). It should be noted, however, that Munhall et al. (1996) has argued against the relationship between stimulus duration and window size, showing comparable windows for audiovisual speech stimuli produced in fast (300 ms), normal (400 ms) and clear (500 ms) speaking conditions. However, it seems unlikely that window size would rapidly morph (during an experiment) after only limited exposure to exaggerated (elongated) speech productions. Rather, it is likely that the process of temporal window contraction is based on cumulative experiences with typical speech in early life which mediate changes in the underlying neural networks engaged during audiovisual speech processing.

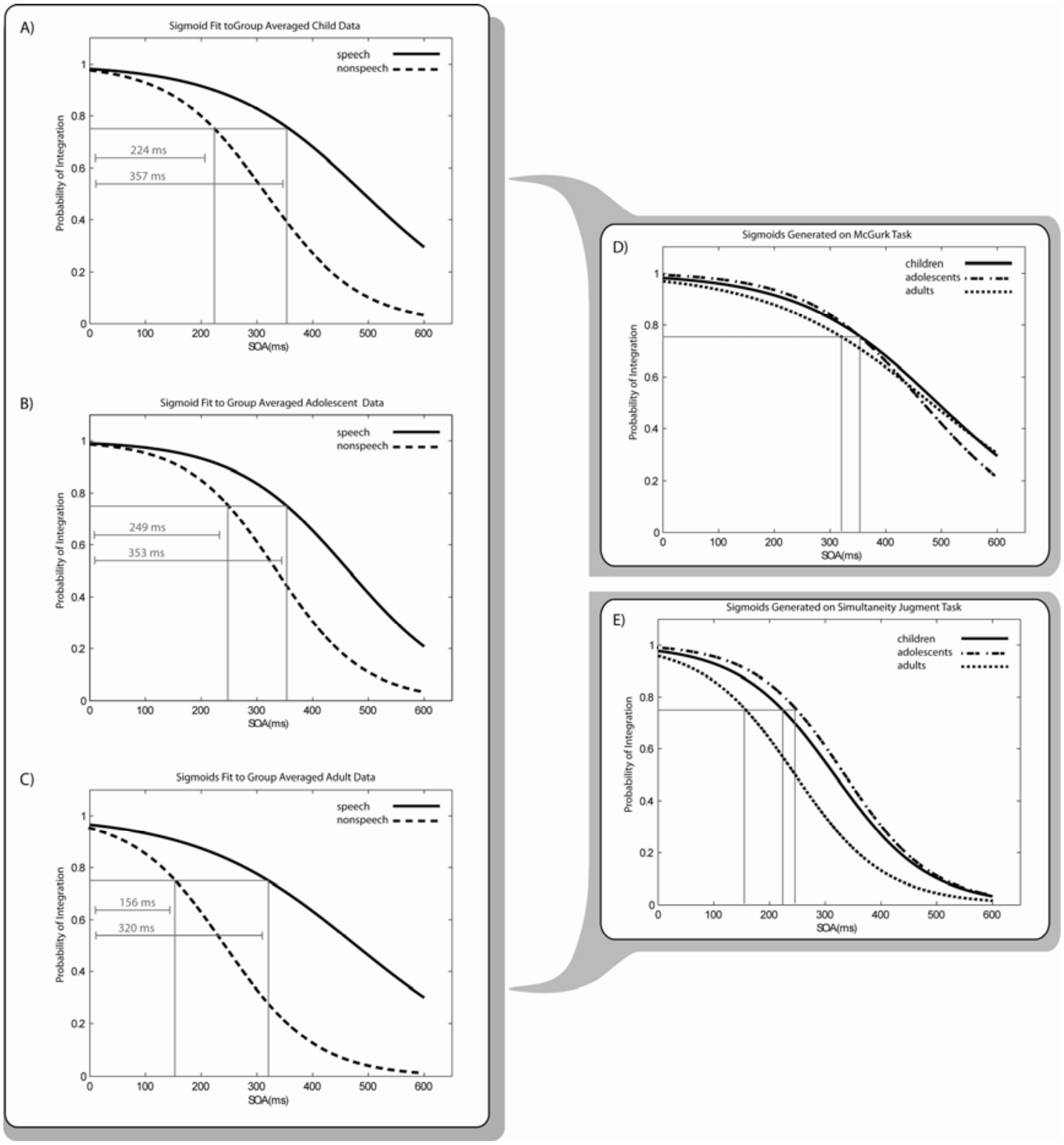


Figure 5.3. Groups show larger windows for speech than nonspeech stimuli. *Comparisons of windows derived to speech and nonspeech stimuli in **A)** children, **B)** adolescents, and **C)** adults. All groups show larger window sizes for speech, but discrepancies are less robust in children and adolescents relative to adults. Comparisons of window sizes across groups for **D)** speech and **E)** nonspeech stimuli reiterate that group differences are exclusively observed on the nonspeech task.*

"Is Speech Special?"

The Cortical Circuits Mediating Speech and Nonspeech Processing Differ

Differences in the developmental period during which mature multisensory temporal processing emerges for speech and nonspeech stimuli may be due to the biological predisposition of the brain for language, or differences in the network (and maturational timeline) of structures recruited to process these stimuli. Neuroimaging work in adults has provided evidence of differential activation for speech and nonspeech signals. Palva et al. observed altered gamma band responses to auditory speech relative to nonspeech stimuli (Palva et al., 2002), and Vouloumanos and colleagues (2001) uncovered increased responses in the middle temporal gyri, left STG and right inferior frontal gyrus to auditory speech compared to tones and complex nonspeech sounds (Vouloumanos, Kiehl, Werker & Liddle, 2001). Others have reported increased activation in the STS to auditory speech relative to tones, and a general increase in the spatial expanse of activation for speech as compared to tones and noise (Binder et al., 2000). A recent review suggests that these and other findings support the existence of a gradient in temporal areas wherein activity is more lateral for speech than basic stimuli (Rauschecker & Scott, 2009).

While the literature characterizing differences in auditory speech and nonspeech processing is quite substantial and only a small sampling of studies are reported here, very little is known about how stimulus differences alter neural activation patterns in the multisensory realm. One recent study has attempted to elucidate such differences by comparing the distribution of activation to audiovisual recordings of manual tools and speech (Stevenson & James, 2009). This contrast indicated that the activation in STG was more anterior for speech when compared to the nonspeech stimuli. Future work is needed to further elaborate such differences. Nonetheless, both auditory and multisensory studies present convincing evidence that the patterns of brain activation associated with the processing of speech and nonspeech stimuli differ, providing a basis for speculations that differences the maturational timelines of these structures may contribute to discrepancies in the realization of adult-like multisensory temporal processing.

Given that the rate of maturation varies for different brain regions (de Haan & Johnson, 2003; Lenroot & Giedd, 2007), earlier contraction of the temporal window for speech may be related to faster development and stabilization of the cortical circuits mediating speech processing. While this interpretation runs counter to traditional views that higher brain regions mature later than low-level cortical domains, evidence is available to support earlier maturation of speech processing networks. For example, in an event-related potential (ERP) study involving 3 - 16 year-olds and adults, adult-like N1c responses to auditory speech stimuli were seen in 7-8 year olds, but this same component was absent in response to tonal stimuli until age 11-12 (Pang et al., 2000). Furthermore, although present by 12 years, the amplitude of the N1c response to tones continued to increase into adulthood, suggesting that the maturation of non-speech circuits is quite delayed. The authors speculate that learning to distinguish speech contrasts may precede the acquisition and refinement of tonal discrimination abilities due to the increased ethological importance of speech. This finding could be extended to the multisensory temporal realm, where we have posited that early, rich language experience drives rapid increases in language proficiency and early window consolidation relative to arbitrary basic stimulus pairs.

Human Infants Preferentially Process Speech Stimuli

Indeed, language is a vital aspect of communication, and it has been suggested that vast gains in receptive and expressive speech ability in early life may be due to the increased salience of speech over other signals. The human bias for language may be rooted in enhanced attention and memory for speech constructs (Jusczyk & Bertoncini, 1988). Support for this notion comes from behavioral work demonstrating a preference in infants for speech over other basic and complex nonspeech signals (Colombo & Bundy, 1981; Friedlander & Wisdom, 1971; Glenn, Cunningham & Joyce, 1981; Vouloumanos & Werker, 2004). For example, Vouloumanos & Werker showed that 2-7 month old infants listened longer to speech than to spectrally and temporally modulated complex nonspeech sounds (Vouloumanos & Werker, 2004). Moreover, nine month olds preferentially responded to a female singing a nursery rhyme over the melody being played on an instrument (Glenn et al., 1981). Thus, it appears that language is more likely to engage attentional processes than other signals, which may segue into more rapid learning of multisensory relationships for speech.

Another critical factor believed to play a role in the development of multisensory temporal processing is experience. Whereas infants and children have very limited exposure to the arbitrary pairings of flashes and tone pips used in the experiments in Chapters II and III, exposure to speech is generally plentiful from birth, enabling dramatic gains in language proficiency in early life. Infants as young as eight months can successfully segment words from continuous speech based on the statistical probability that adjacent sounds will appear together (Saffran, Aslin & Newport, 1996), and by 10-12 months canalization studies show that infants become tuned to the phonetic constructs in their native language (Werker & Tees, 1984). Such increases in native language competencies are also reflected in growth of expressive speech abilities. Typically developing children progress from producing one word at around 12 months to producing approximately 50 words at 18 months and to forming simple sentences by three years of age (Kuhl, 2004). Regular exposure to speech is linked with rapid growth in language abilities and such changes may also enable faster window consolidation for speech.

An Experience-Based Model of Multisensory Temporal Window Contraction

Single unit physiology studies in animals have indicated that postnatal experience is necessary for the establishment of adult-like responses to multisensory stimuli (Wallace, Carriere, Perrault, Vaughan & Stein, 2006; Wallace & Stein, 1997; Wallace & Stein, 2001). Moreover, infant studies show that babies can detect simple amodal temporal stimulus properties in early life (e.g., synchrony), but that the capacity to identify more sophisticated temporal cues (e.g., rate) emerges later, presumably building upon previously acquired multisensory skills (Lewkowicz, 2000). Whereas infants have ample opportunities to learn the properties of speech and language from birth, exposure to basic stimulus constructs is far more limited. We suspect that this discrepancy in experience may be the basis for differences in the timeline to maturation of audiovisual temporal processing.

Results of the Simultaneity Judgment studies contained in this thesis show age-dependent decreases in temporal window size for nonspeech stimuli. No such effect was observed on the McGurk task, presumably because window consolidation was complete by 6 years of age for speech, although the possibility of a fixed window size from birth cannot be ruled out. To our knowledge, no evidence is currently available regarding window sizes to speech in infants. Operating under the assumption that our lower age limit was inadequate for identifying maturational changes, we have put forth a model to explain experience based effects on window contraction (Figure 5.4.) The model prediction holds that an unspecified amount of cumulative experience enables window consolidation, and that such experience is acquired faster for speech resulting in realization of adult-like processing sometime between birth and 6 years of age. Because exposure to basic stimulus pairs is less regular, the timeline for consolidation is extended, with window size stabilizing in middle childhood to adolescence. The time at which adult-like processing is realized represents the closure of the sensitive period for window contraction. As such additional experience does not further decrease window size for speech or nonspeech stimuli, unless strategically altered as with training. Powers et al. (2009) reported lasting changes in the temporal binding windows of mature adults engendered via feedback-mediated training. Repeated, focused exposure to strategic audiovisual pairings altered multisensory temporal performance and presumably the internal representation of multisensory stimulus relations, although the extent of such changes is limited. Hence, early experience with audiovisual pairings is critical to the establishment of proper multisensory temporal processing abilities.

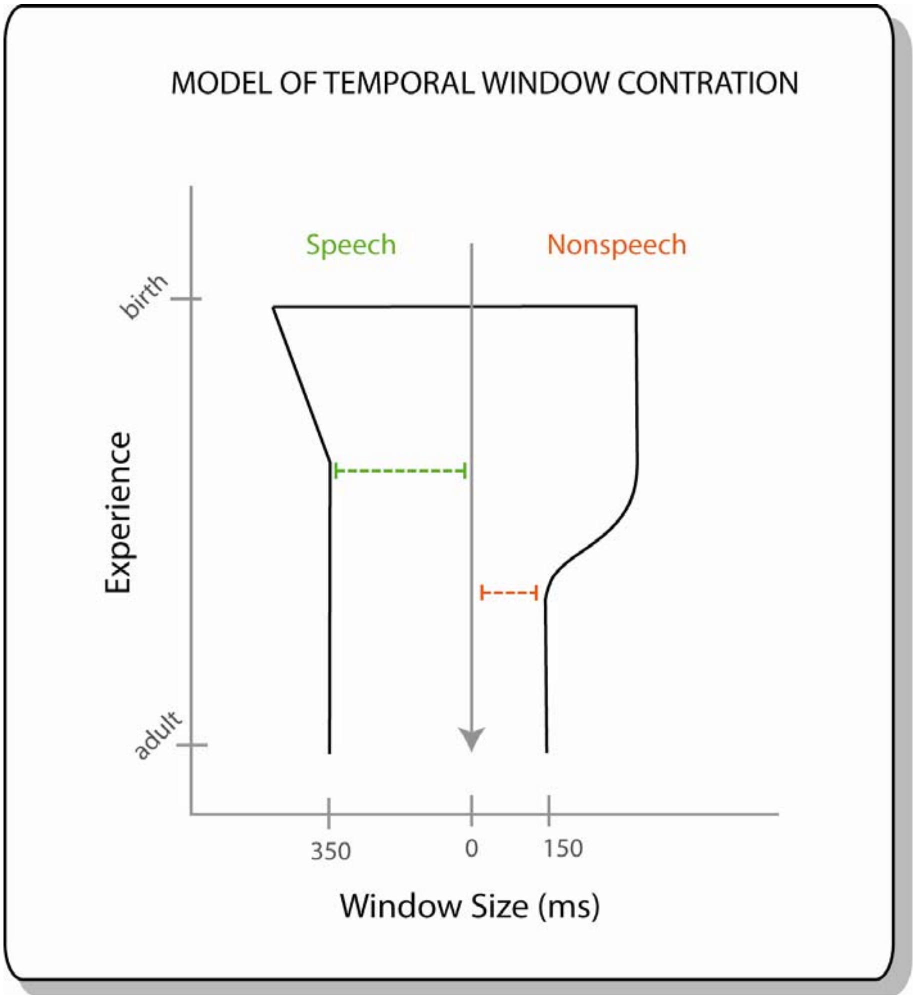


Figure 5.4. An experience-based model accounts for faster contraction of the temporal binding window for speech relative to nonspeech stimuli.

Manifestations of Early Sensory Deprivation on Unisensory and Multisensory Processing-Critical Periods for Development

Repercussions of Hearing and Vision Impairment in Humans

Studies involving individuals with hearing and vision impairment reinforce the importance of experience in mediating the development of multisensory integration abilities and the role of the sensory environment in shaping unisensory and multisensory percepts. These studies indicate that sensory deprivation during development manifests itself in atypical patterns of neural organization and can have lasting consequences on behavior and perception. For example, neuroimaging work shows activity in the primary and secondary visual cortices in individuals with early onset blindness during tactile stimulation or when reading Braille (Amedi, Raz, Pianka, Malach & Zohary, 2003; Obretenova, Halko, Plow, Pascual-Leone & Merabet, 2009; Sadato, Okada, Honda & Yonekura, 2002; Sadato et al., 1996), which is not seen in individuals with normal vision or late onset blindness (> 14- 16 years) (Cohen et al., 1999; Sadato et al., 2002; Sadato et al., 1996). Such a result illustrates that regions of cortex devoid of their normal inputs can be "recruited" by other sensory modalities, and that this compensatory plasticity may form the basis for some of the enhanced capacities reported in these individuals. For example, performance of congenital and early blinded individuals on a vibrotactile discrimination tasks is superior to that of late blinded subjects (14 years) and sighted participants (Boven, Hamilton, Kauffman, Keenan & Pascual-Leone, 2000; Wan, Wood, Reutens & Wilson). Behavioral enhancements are not restricted to the somatosensory realm; rather, individuals with early blindness exhibit improved pitch discrimination (Gougoux et al., 2004; Wan, Wood, Reutens & Wilson, 2010b), auditory localization abilities (Lessard, Pare, Lepore & Lassonde, 1998; Roder et al., 1999), and superior memory for environmental sounds (Röder & Rösler, 2003), speech

(Röder, Rösler & Neville, 2001) and speechlike noise (Amedi et al., 2003), but also see {Röder, 2003 #748} for a discussion of some inconsistencies in such findings.

The age of onset of blindness appears to a critical factor in determining the extent of alterations in cortical organization and behavioral enhancements in non-visual modalities. Whereas subjects with congenital or early onset blindness exhibit profound compensatory plasticity (e.g., enhanced tactile and auditory processing), individuals that lose their vision around 14-16 years of age or later demonstrate patterns of activation and behavior more similar to those of sighted participants (Cohen et al., 1999; Sadato et al., 2002)..

The concept of a sensitive period appears to be a domain general phenomenon, having also been reported in the auditory system, where the age of onset of hearing loss and duration of deafness is strongly correlated with the development of auditory communication abilities. Comparison of receptive and expressive language ability in early- (by 6 months) and later-identified (after 6 months) deaf and hard-of-hearing children indicated that early identification (i.e., diagnosis of hearing loss) and intervention (e.g., fitting of amplification) was consistent with significantly improved language abilities at one to three years of age (Yoshinaga-Itano, Sedey, Coulter & Mehl, 1998). Language abilities were universally poor among later-identified groups (i.e., 7-12 months, 13-18 months, 19-24 months and 25-34 months). Similar findings have been reported in children receiving cochlear implants. Children implanted by three and a half years of age exhibited P1 cortical auditory evoked potentials (CAEPs) that were similar to those of normal hearing controls within six to eight months of implant activation (Sharma, Dorman & Spahr, 2002a, 2002b), despite the fact that the implant fails to replicate the normal patterns of auditory stimulation (and activation). In contrast, the majority of subjects implanted after age seven failed to exhibit response normalization (Sharma et al., 2002b). Taken together, findings from electrophysiological and behavioral studies in children and adults suggest that alterations in experience during the critical period for development produce lifelong changes in unisensory function; changes that have strong ramifications for multisensory processing.

While few studies have investigated the impact of early onset hearing and vision loss on multisensory processing, those that are available complement the unisensory literature and suggest that critical periods govern the development of the multisensory systems as well. Gilley et al (2010) evaluated reaction times of early- and late-implanted children and subjects with normal hearing on an auditory, visual, and audiovisual detection task. The multisensory responses of early-implanted and normal hearing subjects violated Miller's Race Model (i.e., reaction times for bimodal presentations exceeded those predicted by the fastest unisensory response), suggesting that auditory and visual information was integrated to produce response times shorter than predicted by additive models (Gilley, Sharma, Mitchell & Dorman, 2010). This effect was not observed in late-implanted subjects. Complementary findings were reported by Schorr and colleagues in five- to 14-year old normal hearing and cochlear implant subjects using a McGurk task (Schorr, Fox, van Wassenhove & Knudsen, 2005). Here, all of the deaf children comprising the good integrator group (i.e., those perceiving a high percentage of fusions on incongruent trials) were implanted prior to 30 months of age, suggesting that early auditory experience through the cochlear implant is necessary for successful development of audiovisual integration of speech and nonspeech stimuli.

Overall findings suggest that early hearing and vision loss can impede the capacity to develop good crossmodal integration abilities and to maximally benefit from multisensory enhancements that provide a competitive advantage in many everyday scenarios. These studies provide a theoretical basis for future investigations into how congenital bilateral hearing loss and early intervention affects multisensory processing, ideas that will elaborated upon in further detail in the *Future Directions* section to follow.

Future Directions

Evaluation of Multisensory Processing in Babies and Children with Hearing Impairment

While studies have provided evidence that earlier implantation increases the likelihood that children will be better integrators of multisensory information, the time course for emergence and maturation of crossmodal abilities has not been systematically studied in this population (at periodic intervals after implantation or hearing aid fitting). Moreover, it remains to be seen whether differences in communication strategies adopted during early life (e.g., auditory-verbal versus combined modality approaches) could produce corresponding changes in multisensory integration.

While investigation of multisensory temporal processing in the hearing impaired population is a natural extension of the current work, there are also a number of interesting questions regarding processing on non-temporal tasks that have yet to be explored. Families with babies and young children with hearing loss adopt distinct communication strategies and therapy techniques (e.g., auditory verbal, auditory oral, total communication, cued or signed speech) to facilitate expressive and receptive speech and language abilities. Given that the environment and exposure to speech structures influences language development, it is plausible that alterations in the input technique may be reflected in differences in the magnitude and frequency of audiovisual integration. For example, the auditory-verbal approach is founded on a learning through listening premise, where the availability of visual cues is restricted (eliminated) during therapy. In contrast, the auditory-oral approach involves the combined use of auditory and visual information (lipreading), and total communication uses audiovisual speech cues and accompanying hand gestures for communication. It is conceivable that differences in early training styles could affect alterations in audiovisual processing. It is predicted that children trained to listen using auditory cues exclusively might exhibit reduced audiovisual integration relative to hearing impaired peers employing multisensory listening strategies. It is of interest to explore whether such a difference exists and whether it has lasting implications on their capacity to benefit from multisensory enhancements.

Potential differences in the likelihood of multisensory integration within the hearing impaired population may also been observed when comparing groups with differing degrees of hearing loss. Based on the principle of inverse effectiveness (maximal integration is produced by weakly effective unisensory stimuli) it might be inferred that audiovisual integration is increased in individuals with greater degrees of hearing loss, where the auditory signal is further degraded. The notion of critical periods for intervention are also worthy of additional attention. Schorr and others have reported that early implantation is necessary for audiovisual fusion and realization of multisensory facilitation (Gilley et al., 2010; Schorr et al., 2005), but the relationship between critical periods and development of multisensory processing abilities has not been evaluated in children fitted with tradition amplification (i.e., hearing aids). Furthermore, the timeline to acquisition of multisensory gains has not been studied. Measurement of audiovisual integration at regular intervals pre- and post-implantation/hearing aid fitting is necessary to uncover the trajectory to development and normalization of multisensory processing. This research could provide critical insight into how the time of intervention and the chosen communication strategy affects realization of multisensory interactions that parlay into robust communication gains.

Assessing the Efficacy of Multisensory-Minded Classroom Modifications on Academic Performance

A recurring theme throughout this thesis is the profound impact of multisensory integration on behavior and perception. Bearing this in mind, how might the classroom environment and academic instruction be optimized to capitalize gains that can be conferred by multisensory integration?

Given that improvements in thresholds for detecting sounds and speech are observed until and perhaps beyond 12 years of age (Jerger & Jerger, 1982; Maxon & Hochberg, 1982; Trehub, Schneider, Morrongiello & Thorpe, 1988), and that children exhibit increased difficulty understanding speech in noise (Buss, Hall, Grose & Dev, 1999; Hall, Grose, Buss & Dev, 2002; Jerger & Jerger, 1982) , strategies aimed at maximizing audiovisual speech benefits may prove especially useful in promoting understanding by increasing signal salience. The classroom configuration is critical to facilitating visual enhancement of auditory speech. All students should be situated such that they can easily view the teacher during lectures. Hollow square or pod configurations should be avoided as all children cannot simultaneously access lipreading cues given that some desks will inevitably face away from the teacher Figure (5.5). Moreover, the traditional set-up could be improved upon by staggering desks to ensure that the view of the teacher is not obstructed for children seated behind the first row. Another critical but often under looked issue relates to note-taking. When writing, children look down at their paper and away from the teacher, causing them to be more vulnerable to communication breakdowns caused by chatter and room noise. Hand-outs are especially useful limiting the need for note-taking and thus reducing potential lapses in understanding. Lastly, do lessons engage a variety of the senses? Whenever possible, teachers should try to engage multiple senses during instruction given that sensory redundancy promotes learning. For example, when learning about shapes given the opportunity to both view and handle blocks could improve one's ability to adequately encode the representations. While basic, these strategic classroom and instruction modifications could substantially promote understanding in children and potentially parlay into improved academic performance. Given the findings in the current theses indicating that changes in audiovisual integration of speech and multisensory temporal processing of nonspeech cues continue into middle childhood and adolescence, such strategies are particularly important for students of all ages who are less capable of drawing on multisensory integrative effects to improve processing of speech and environmental cues.

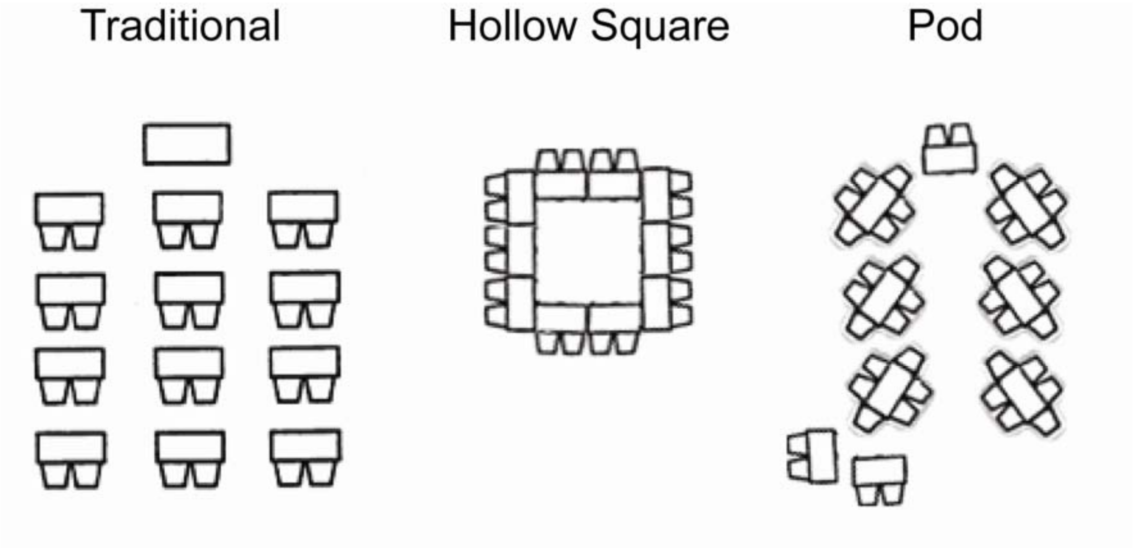


Figure 5.5. Optimal classroom set-ups require that all children can see the teacher during lecture. *For lecture based instruction the traditional set-up is preferable to the hollow square and pod configurations since desks (rectangles) and chairs (trapezoids) are also facing the same direction. If the teacher stands in front, all children can adequately access visual cues to facilitate understanding.*

References

Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early /`visual/' cortex activation correlates with superior verbal memory performance in the blind. *Nat Neurosci, 6*(7), 758-766.

Arai, T., & Greenberg, S. (1997). *The temporal properties of spoken Japanese are similar to those of English.* Paper presented at the Eurospeech, Rhodes, Greece.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci, 30*(7), 2414-2417.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cereb. Cortex, 10*(5), 512-528.

Boven, R. W. V., Hamilton, R. H., Kauffman, T., Keenan, J. P., & Pascual-Leone, A. (2000). Tactile spatial resolution in blind Braille readers. *Neurology, 54*(12), 2230-2236.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220.

Buss, E., Hall, J. W., III, Grose, J. H., & Dev, M. B. (1999). Development of Adult-Like Performance in Backward, Simultaneous, and Forward Masking. *J Speech Lang Hear Res, 42*(4), 844-849.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14*(17), 2213-2218.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649-657.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 1001-1010.

Cohen, L. G., Weeks, R. A., Sadato, N., Celnik, P., Ishii, K., & Hallett, M. (1999). Period of susceptibility for cross-modal plasticity in the blind. *Annals of Neurology, 45*(4), 451-460.

Colombo, J., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development, 4*, 219-223.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am, 119*(6), 4065-4073.

de Haan, M., & Johnson, M. H. (2003). Mechanisms and theories of brain development. In M. de Haan & M. H. Johnson (Eds.), *The Cognitive Neuroscience of Development*. New York: Psychology Press.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Friedlander, B. Z., & Wisdom, S. S. (1971). *Pre-verbal infants' selective operant responses for different levels of auditory complexity and language redundancy*. Paper presented at the Annual Meeting of the Eastern Psychological Association.

Gilley, P. M., Sharma, A., Mitchell, T. V., & Dorman, M. F. (2010). The influence of a sensitive period for auditory-visual integration in children with cochlear implants. *Restor Neurol Neurosci, 28*(2), 207-218.

Glenn, S. M., Cunningham, C. C., & Joyce, P. F. (1981). A Study of Auditory Preferences in Nonhandicapped Infants and Infants with down's Syndrome. *Child Development, 52*(4), 1303-1307.

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., & Belin, P. (2004). Neuropsychology: Pitch discrimination in the early blind. *Nature, 430*(6997), 309-309.

Hall, J. W. I., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear and Hearing, 23*(2), 159-165.

Jerger, S., & Jerger, J. (1982). Pediatric Speech Intelligibility Test: Performance-Intensity Characteristics. *Ear and Hearing, 3*(6), 325-334.

Jones, J. A. C. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport, 14*(8), 1129-1133.

Jusczyk, P. W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Lang Speech, 31 (Pt 3)*, 217-238.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat Rev Neurosci, 5*(11), 831-843.

Lenroot, R. K., & Giedd, J. N. (2007). The structural development of the human brain as measured longitudinally with magnetic resonance imaging. In D. Coch, K. W. Fischer & G. Dawson (Eds.), *Human Behavior, Learning, and the Developing Brain*. New York: The Guilford Press.

Lessard, N., Pare, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature, 395*(6699), 278-280.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull, 126*(2), 281-308.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb Cortex, 15*(10), 1621-1631.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: sensitivity and discrimination. *Ear Hear, 3*(6), 301-308.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Munhall, K. G., & Tohkura, Y. (1998). Audiovisual gating and the time course of speech perception. *The Journal of the Acoustical Society of America, 104*(1), 530-539.

Obretenova, S., Halko, M. A., Plow, E. B., Pascual-Leone, A., & Merabet, L. B. (2009). Neuroplasticity associated with tactile language communication in a deaf-blind subject. [Original Research Article]. *Frontiers in Human Neuroscience, 3*.

Palva, S., Palva, J. M., Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Kaila, K., et al. (2002). Distinct Gamma-Band Evoked Responses to Speech and Non-Speech Sounds in Humans. *J. Neurosci.*, 20026134.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci, 12*(6), 718-724.

Röder, B., & Rösler, F. (2003). Memory for environmental sounds in sighted, congenitally blind and late blind adults: evidence for cross-modal compensation. *International Journal of Psychophysiology, 50*(1-2), 27-39.

Röder, B., Rösler, F., & Neville, H. J. (2001). Auditory memory in congenitally blind adults: a behavioral-electrophysiological investigation. *Cognitive Brain Research, 11*(2), 289-303.

Roder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature, 400*(6740), 162-166.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Percept Psychophys, 59*(3), 347-357.

Sadato, N., Okada, T., Honda, M., & Yonekura, Y. (2002). Critical Period for Cross-Modal Plasticity in Blind Humans: A Functional MRI Study. *NeuroImage, 16*(2), 389-400.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380*(6574), 526-528.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science, 274*(5294), 1926-1928.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication, 26*(1-2), 75-87.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proc Natl Acad Sci U S A, 102*(51), 18748-18750.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123 Pt 12*, 2400-2406.

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

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002a). Rapid development of cortical auditory evoked potentials after early cochlear implantation. *Neuroreport, 13*(10), 1365-1368.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002b). A Sensitive Period for the Development of the Central Auditory System in Children with Cochlear Implants: Implications for Age of Implantation. *Ear and Hearing, 23*(6), 532-539.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *J Exp Psychol Hum Percept Perform, 35*(2), 580-587.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage, 44*(3), 1210-1223.

Trehub, S. E., Schneider, B. A., Morrongiello, B. A., & Thorpe, L. A. (1988). Auditory sensitivity in school-age children. *Journal of Experimental Child Psychology, 46*(2), 273-285.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PLoS One, 2*(1), e742.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience, 13*(7), 994-1005.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science, 7*(3), 270-276.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Jr., Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *J Neurosci, 26*(46), 11844-11849.

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci, 17*(7), 2429-2444.

Wallace, M. T., & Stein, B. E. (2001). Sensory and Multisensory Responses in the Newborn Monkey Superior Colliculus. *J. Neurosci., 21*(22), 8886-8894.

Wan, C. Y., Wood, A. G., Reutens, D. C., & Wilson, S. J. (2010a). Congenital blindness leads to enhanced vibrotactile perception. *Neuropsychologia, 48*(2), 631-635.

Wan, C. Y., Wood, A. G., Reutens, D. C., & Wilson, S. J. (2010b). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia, 48*(1), 344-348.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63.

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998). Language of Early- and Later-identified Children With Hearing Loss. *Pediatrics, 102*(5), 1161-1171.