The development of multisensory speech perception continues into the late childhood years

Lars A. Ross[1,2,3,#], **Sophie Molholm**[1,2,3], **Daniella Blanco**[1,3], **Manuel Gomez-Ramirez**[2,3], **Dave Saint-Amour**[2,4,5], and **John J. Foxe**[1,2,3,#]

[1] The Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center (CERC), Departments of Pediatrics and Neuroscience, Albert Einstein College of Medicine, Van Etten Building – Wing 1C, 1225 Morris Park Avenue, Bronx, New York 10461, USA

[2] The Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, New York 10962, USA

[3] Program in Cognitive Neuroscience, Departments of Psychology & Biology, City College of the City University of New York, 138th Street & Convent Avenue, New York, New York 10031, USA

[4] Centre de recherche, CHU Sainte-Justine, 3175, Côte-Sainte-Catherine Montréal, Québec, H3T 1C5, Canada

[5] Département de psychologie, Université du Québec à Montréal (UQAM), C.P. 8888 succursale Centre-ville, Montréal, Québec, H3C 3P8, Canada

Abstract

Observing a speaker's articulations substantially improves intelligibility of spoken speech, especially under noisy listening conditions. This multisensory integration of speech inputs is crucial to effective communication. Appropriate development of this ability has major implications for children in classroom and social settings, and deficits in it have been linked to a number of neurodevelopmental disorders, especially autism. It is clear from structural imaging studies that there is a prolonged maturational course within regions of the perisylvian cortex that persists into late childhood, and these regions have been firmly established as crucial to speech-language functions. Given this protracted maturational timeframe, we reasoned that multisensory speech processing might well show a similarly protracted developmental course. Previous work in adults has shown that audiovisual enhancement in word recognition is most apparent within a restricted range of signal-to-noise ratios. Here we asked when these properties emerge during childhood by testing multisensory speech recognition abilities in typically developing children aged between 5 and 14, comparing them to adults. By parametrically varying signal-to-noise ratios (SNRs), we found that children benefited significantly less from observing visual articulations, displaying a considerably less audiovisual enhancement. The findings suggest that improvement in the ability to recognize speech-in-noise and in audiovisual integration during speech perception continues quite late into the childhood years. The implication is that a considerable amount of multisensory learning remains to be achieved during the later schooling years and that explicit efforts to accommodate this learning may well be warranted.

Keywords

[#]Correspondence: john.foxe@einstein.yu.edu or lars.ross@einstein.yu.edu.

Introduction

It is well established that viewing a speaker's articulatory movements can substantially enhance the perception of auditory speech, especially under noisy listening conditions (Sumby and Pollack, 1954; Erber, 1969, 1971, 1975; Ross et al. 2007a, b). The magnitude of this audiovisual gain depends greatly on the relative fidelity of the auditory speech signal itself, particularly on the signal-to-noise ratio (SNR), and it has been suggested in the past that this gain increases with decreasing SNR ratios (Sumby and Pollack, 1954, Erber, 1969, 1971, 1975, Callan et al., 2003). However, recent evidence in healthy adults from studies investigating audiovisual enhancement of the perception of words suggests that this gain tends to be largest at "intermediate" SNRs; that is, between conditions where the auditory signal is almost perfectly audible and where it is completely unintelligible (Ross et al., 2007a,b; Ma et al., 2009).

Sensitivity to coordinated audiovisual speech inputs manifests remarkably early in development, considerably before the acquisition of language. Evidence for this has been established using an audiovisual matching technique in which infants are presented with videos of speakers producing either congruent speech or speech where the visual articulation is not matched to the sound. Children's preference is measured as a function of the amount of time spent fixating a given stimulus display (Burnham and Dodd, 1998; Dodd, 1979; Kuhl and Meltzoff, 1982, 1984; Patterson and Werker, 2003) or by the amplitude of sucking (Walton and Bower, 1993). Preference for congruent audiovisual speech has been found in 2-month old (Patterson and Werker, 2003) and 4-month old infants for native vowels (Kuhl and Meltzoff, 1982, 1984), and in 6-month old children for simple syllables (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Two-month old infants also show preference for a talker producing congruent over incongruent ongoing speech (Burnham and Dodd, 1998; Dodd, 1979). There is even evidence for audiovisual matching of speech in newborns (Aldridge, Braga, Walton, & Bower, 1999). Despite the early appearance of the preference for congruent audiovisual speech, there is also ample evidence for developmental change through experience and maturation.

It is known from studies using so-called McGurk-type tasks that there are age-related differences in the susceptibility to visual speech (Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986, McGurk and MacDonald, 1976; Sekiyama and Burnham, 2008). The McGurk effect is a rather remarkable multisensory illusion whereby dubbing a phoneme onto an incongruent visual articulatory speech movement can lead to an illusory change in the auditory percept (McGurk and MacDonald, 1976; Saint-Amour, DeSanctis, Molholm, Ritter, & Foxe, 2007). In their original study, McGurk and MacDonald (1976) reported that children of 3 to 5 years and 7 to 8 years showed less susceptibility to the influence of incongruent visual speech than adults. This finding was later confirmed in a series of experiments investigating visual influence on the identification of auditory syllables ranging on a continuum between /ba/ and /da/ (Massaro, 1984; Massaro et al., 1986). It was consistently shown that children of 4 to 6 and 6 to 10 years of age were less influenced by the visual articulation of an animated speaker than adults. Similarly, Hockley and Polka (1994) reported a gradual developmental increase in the influence of visual articulation across the ages of 5, 7, 9, and 11 years. More recently, Sekiyama and Burnham (2008) showed that the audiovisual integration of speech, also indexed by susceptibility to the McGurk illusion, sharply increased over the age-range from 6 to 8 years in native English speakers.

The ability to recover unisensory auditory speech when it is masked in noise has also been shown to increase with advancing age. In the classroom environment, younger children are more distracted by noise than older children (Hetu, Truchon-Gagnon, & Bilodeau, 1990) and

are less likely to identify the last word in a sentence that is presented in multi-talker babble (Elliott, 1979; Elliott, Connors, Kille, Levin, Ball, and Katz, 1979), which is also the case for words and sentences presented in spectral noise (Nittrouer and Boothroyd, 1990). These deficits have been attributed to a variety of factors including utilization of sensory information, and linguistic and cognitive developmental factors (e.g. Eisenberg, Shannon, Schaefer Martinez, Wygonski, and Boothroyd, 2000; Fallon, Trehub, and Schneider, 2000).

Recent advances in neuroimaging technology have brought new insights into the neurophysiological changes that accompany the maturation of cognitive functions. It has been shown that cortical anatomy in perisylvian language areas shows a relatively long developmental trajectory with a relatively protracted grey matter thickening (Sowell et al., 2004). It is a reasonable assertion that this long maturation course is associated with the long duration of language development and the fine tuning of language skills. The increase in formal language learning in the early school years is associated with a sharp increase in face-to-face communication in a typical classroom setting. Even though developmental changes in cortical regions underlying more basic sensory and perceptual functions are thought to terminate earlier than in perisylvian regions (Shaw et al., 2008), it is quite possible that neural structures underlying the integration of auditory and visual speech develop in parallel with higher order language functions late into adolescence.

The reported studies show convincingly that the influence of visualized articulations, while certainly present, is clearly also weaker in infants and that it develops across childhood until it reaches adult levels. The aforementioned studies of audiovisual integration in speech perception have used intact speech signals and it is therefore not known whether this weaker visual effect is uniform over a range of SNR levels. As mentioned before, visual enhancement of word recognition is highly dependent on the quality of the speech signal and our previous work has shown that audiovisual benefit in word recognition follows a characteristic pattern in adulthood (Ross et al., 2007a,b). We have hypothesized that this pattern must emerge across childhood as acoustic, linguistic and articulatory skills continue to develop rapidly (see Saffran, Werker and Werner, 2006 for a review) and exposure to different lexical environments is encountered. It is possible that full maturity of this system may even be delayed until adolescence. Here, we asked if and when these properties of audiovisual speech integration emerge during childhood by testing multisensory speech recognition abilities in a cohort of typically developing children and adolescents across an age-range from 5 to 14 years, and comparing their performance to a cohort of healthy teenagers and adults (16–46 years of age).

A set of simple predictions was made. As with previous work by others, we expected that recognition of words presented in noise would improve with age (Eisenberg et al., 2000, Elliott, 1979; Elliott et al., 1979; Massaro et al., 1986), with the youngest children showing the lowest recognition scores. We have shown that when monosyllabic words are embedded in various levels of noise the maximal audiovisual gain occurs at approximately 20% auditory-alone performance (at −12dB SNR in adults) and so we expected that children would also show maximal benefit where their auditory-alone performance was at about 20% word-recognition. However, since this level of performance is expected to shift with increasing age during childhood to lower SNRs, it stands to reason that maximal benefit will not stabilize until auditory-alone performance nears full maturity, and as such, we predicted that overall multisensory gain would be considerably lower in children, especially at younger ages. Note that this latter prediction is not meant to imply that multisensory benefit "follows" unisensory development, since such a serial learning process is highly improbable. Rather, we would hold that the ordered development of unisensory auditory recognition processes is just as reliant on intact multisensory learning processes as *vice versa*. We were especially interested in establishing whether there were specific age-brackets during which

multisensory speech processes showed a particularly steep developmental trajectory and to establish the age-bracket at which the emergence of a fully adult pattern would be observed.

Methods

Participants

44 typically developing children (range: 5 to 14 years; $M = 9.59$; $SD = 2.68$) and 14 neurotypical adults (range: 16 to 56 years; $M = 32.36$; $SD = 12.43$) participated in this study. Typical Development was defined here according to the following criteria, as ascertained through parent interview: 1) No history of neurological, psychological or psychiatric disorders, 2) No history of head trauma or loss of consciousness, 3) No current or past history of psychotropic medication use, 4) Age appropriate grade-performance. All participants were native English speakers, had normal or corrected to normal vision and normal hearing. Two of the children were bilingual (Spanish and Chinese respectively) but in both cases English was acquired early as the primary language. Comparison of the performance data of these children with a sample of children of similar age revealed no differences. Informed consent was obtained from all adults, children and their caretakers. All procedures were approved by the Institutional Review Board of the City College of the City University New York and by the Institutional Review Board of the Albert Einstein College of Medicine, and were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Stimuli and task

Stimulus material consisted of digital recordings of 300 simple monosyllabic words spoken by a female speaker. This set of words was a subset of the stimulus material created for a previous experiment in our laboratory (Ross et al., 2007a,b). These words were taken from the "MRC Psycholinguistic Database" (Coltheart, 1981) and were selected from a well-characterized normed set based on their written-word frequency (Kucera and Francis, 1967). The subset of words for the present experiment is a careful selection of simple, high-frequency words from a child's everyday environment and is likely to be in the lexicon of children in the age-range of our sample.

The recorded movies were digitally re-mastered, so that the length of the movie (1.3 sec) and the onset of the acoustic signal were similar across all words. Average voice onset occurred at 520ms after movie onset (SD= 30ms). Noise onset was the same time as the movie onset, 520ms before the beginning of the speech signal. The words were presented at approximately 50dBA FSPL and six different levels of pink noise were presented simultaneously with the presentation of the words at no noise (NN), 53, 56, 59, 62, and 65dBA FSPL. In one condition the words were presented without additional noise. The signal-to-noise ratios (SNRs) are therefore NN, -3, -6, -9, -12, –15, −18dBA FSPL. These SNRs were chosen to cover a performance range in the auditory-alone condition from 0% recognized words at the lowest SNR to almost perfect recognition performance with no noise.

The movies were presented on a 17-inch portable laptop computer monitor at a distance of approximately 50cm from the participant. The face of the speaker extended approximately 12° of visual angle horizontally and 12.5° vertically (hairline to chin). The words and pink noise were presented over headphones (Sennheiser©, model HD 555).

The main experiment consisted of three conditions presented in randomized order: In the auditory-alone condition (A_{alone}) the auditory recording of the words were presented in conjunction with a still image of the speakers face; in the audiovisual condition (AV) observers saw the face of the female speaker articulating the words. Finally, in the visual

alone condition only the speaker's articulations were seen only with auditory noise but without any auditory speech stimulus. The word stimuli were presented in a fixed order and the condition (the noise level and whether it is presented as A_{alone}, V_{alone} or AV) was assigned to the word in a pseudorandom order. Stimuli were presented in 15 blocks of 20 words each. No words were repeated. Participants were instructed to watch the screen and report which word they heard. The experimenter assured that eye fixation was maintained by reminding participants, if necessary. If a word was not clearly understood they were asked to guess which word was presented. The experimenter was seated at approximately 1 m distance from the participant at a 90° angle to the participant- screen axis. The experimenter recorded whether the response exactly matched the word presented. Any other response was recorded as an incorrect answer.

Analyses

We divided our participants into five age groups (5 to 7: n = 10; 8 to 9: n = 11; 10 to 11: n = 13; 12 to 14: n = 10; 16 to 56: n =14) and submitted percent correct responses to a repeated measures analysis of variance (RM- ANOVA) with factors of stimulus condition (auditory versus audiovisual), SNR level (7 levels), and the between subjects factor of age group (5 groups). We expected significant main effects of condition, SNR level and age group as well as an interaction between condition and SNR level replicating previous findings by Ross et al. (2007a, 2007b) and Ma et al. (2009). We expected developmental change in the ability to benefit from visual speech to manifest itself as an interaction of the group factor with condition and SNR level. To see whether age differences in AV- gain were manifested differently across SNR levels, we conducted a multivariate analysis of variance with factors of group and AV- gain at the four lowest SNRs. This analysis was performed at the four lowest SNRs because the variance at higher SNRs became increasingly constrained by ceiling performance.

Audiovisual enhancement (or AV- gain) was operationalized here as the difference in performance between the AV and the A_{alone} condition (AV – A_{alone}). The issue of ways to characterize AV benefit is somewhat contentious and has been discussed more exhaustively elsewhere (see Holmes, 2007; Ross et al., 2007; Sumby and Pollack, 1954). We therefore limit ourselves here to a brief explanation of our motivation to use simple difference scores as an index of AV- benefit. Different methods of characterizing audiovisual gain have been used in the multisensory literature. If gain is defined as the percentage increase relative to the A_{alone} condition, then AV benefit is exaggerated at the lowest SNRs (Ross et al., 2007a, Holmes 2007). Where performance approaches ceiling levels at high SNRs, AV gain is naturally constrained. A widely used method, initially suggested in their seminal paper by Sumby and Pollack (1954) adjusts AV- gain by the room for improvement left for AV performance which becomes drastically limited with increasing intelligibility. Unfortunately, this approach overcompensates for the ceiling effect resulting in a largely exaggerated benefit at high SNRs. It is, however, not at all intuitive that the largest AV- benefit appears at high levels of auditory intelligibility and stands in stark contrast to what we know about the physiological underpinnings of AV integration (Stein and Meredith, 1993). Using the simple difference score has the advantage that such artifacts are avoided and can be used without concern at low SNRs where improvement is not constrained by a performance ceiling. While we assess A, V and AV performance across a wide range of SNRs, we will limit important aspects of our analyses here to the four lowest SNRs. For a more detailed discussion of the characterization of AV-enhancement in speech perception, see Ross et al. (2007a).

Finally, the V_{alone} condition was compared between groups with independent t-tests and correlated with performance in the AV- condition and also with overall AV- gain. This analysis was conducted to test previous evidence that AV enhancement is related to

speechreading ability (Massaro, 1984; Massaro et al. 1986) although it bears mentioning that others have shown no correlation (e.g. Cienkowski and Carney, 2002; Gagné, Querengesser, Folkeard, Munhall, and Masterson, 1995; Munhall, 2002; Ross et al., 2007b; Watson, Qiu, Chamberlain, and Li, 1996).

Results

The effect of speaker articulation, SNR level and age on recognition performance

The addition of visual articulation reliably enhanced word-recognition performance (Sumby and Pollack, 1954, Ross et al., 2007a, b; Ma et al., 2009), resulting in a significant main effect of stimulus condition [$F(1, 53) = 964. 89$; $p < 0.001$, $\eta^2 = 0.95$]. As expected, there was also a significant main effect of SNR [$F(6, 318) = 1127.97$; $p < 0.001$, $\eta^2 = 0.96$] on recognition performance (Sumby and Pollack, 1954, Ross et al., 2007a, b). A significant interaction between condition and SNR- level $F(1, 53) = 87.66$, $p < 0.001$, $\eta^2 = 0.62$] suggested that the AV-enhancement was dependent on the level of noise.

The five groups differed significantly in overall performance [$F(4, 53) = 14.62$, $p < 0.001$, $\eta^2 = 0.53$]. Figure 1 displays performance in the A_{alone}, V_{alone} and AV conditions and AV-gain for all five groups averaged over the four lowest SNRs, where AV- performance was not limited by a performance ceiling. It reveals that this group- effect was mainly determined by performance differences in the AV condition across age groups. AV-performance increased from 27.8% in the youngest sample (5–7) to 52.7% in adults.

The ability to recognize the words in the A_{alone} condition changed surprisingly little across age groups with an average level of 7.1% in 5–7 years of age and 11.8% in adults. However, the influence of age-group on the overall A_{alone} performance (averaged over all SNR levels) was significant [$F(4, 52) = 8.42$; $p < 0.001$]. Subsequent multivariate general linear model (GLM) revealed that group differences were only apparent at higher SNR levels [−9dBA: $F(4,52) = 3.39$; $p = 0.02$; − 6dBA: $F(4,52) = 3.08$; $p = 0.02$; −3dBA: $F(4,52) = 4.71$; $p < 0.01$; NN: $F(4,52) = 4.26$; $p < 0.01$]. This interesting dynamic in group differences depending on condition was supported by a significant interaction between condition and age-group [$F(4, 53) = 7.24$, $p < 0.001$, $\eta^2 = 0.35$].

There was a significant three-way interaction between condition, SNR-level and age-group [$F(4, 53) = 7.58$, $p < 0.001$, $\eta^2 = 0.36$], whereas the interaction between SNR-level and age-group was not significant [$F(4, 53) = 1.16$, $p = 0.34$, $\eta^2 = 0.08$] reflecting the non-linearity in the increase in VA- performance over age apparent in Fig. 1. While there was a substantial increase in AV performance in our youngest age group 5–7 to the next of 8–9 there was very little difference in gain in the 8–9 and the 10–11 year old groups. In contrast, there was a substantial increase in gain seen in the 12–14 year old group and this approached adult levels of gain.

This observed increase in AV- gain over age did not depend on an increase in the ability to speechread (V_{alone}). Speech reading performance did not increase over age as the bar graph in Figure 1 clearly reveals. A separate one-way ANOVA with group as a factor and V_{alone} performance as a dependent variable confirmed this observation ($F(1, 53) = 0.68$; $p = 0.61$). We also tested whether V_{alone} performance was related (*Pearson's r*) to overall AV- gain at the four lowest SNRs where AV-gain was maximal and not constrained by ceiling effects, and found a near significant relationship ($r = 0.24$; $p = 0.06$). We subsequently tested for possible covariance of speechreading with AV-gain in adults and children (collapsed over all age groups) and found that V_{alone} correlated with AV- gain in children ($r = 0.33$; $p = 0.03$) but not in adults ($r = −0.06$; $p = 0.83$). This relationship is not likely to be due to overall performance differences among individual children because V_{alone} performance did

not covary with performance in the A_{alone} (collapsed over the 4 lowest SNRs) condition ($r =$ 0.056; $p=$ 0.725).

Audiovisual gain and SNR

We replicated findings originally demonstrated by Ross et al. (2007a) showing that the magnitude of audiovisual gain in word recognition was critically dependent on SNR, showing maximal enhancement at "intermediate" SNR levels in adults (−12dB SNR) (see Fig. 2.). At this SNR, about 11% of the words were recognized in the A_{alone} condition. An intermediate maximum was also apparent in our sample of children (see Fig. 2), however the enhancement at each individual SNR was considerably lower than in adults. In the youngest children between the ages of 5–7 AV-gain also increased with increasing SNR but the overall gain is lower and the characteristic peak at −12dB is missing showing an overall shallower gain curve with a maximum at −9dB SNR.

Developmental change in AV- enhancement at different SNRs

We further explored whether AV- performance developed uniformly at all SNR levels. We plotted AV- gain (median1) for the four lowest SNRs across the five age groups (see Fig. 3.). AV- gain at all four SNR levels increased linearly between the youngest age group and adults. This increase was particularly stable at −12dBA SNR and the lowest SNR at − 18 dBA.

This pattern was confirmed by multivariate GLM showing that AV- gain was consistent over all age groups and was manifest at all of the 4 lowest SNRs (see Table 1.).

Discussion

In this study we assessed the benefit conferred by viewing visual articulation on the perception of words embedded in acoustic noise over an age span from 5 years of age into adulthood. We expected that younger children would have more difficulty perceiving speech in masking noise, and that while they would certainly benefit from exposure to visualized speaker articulations, this benefit would be significantly lower than that seen in adults. Of primary interest was to quantify the enhancement of multisensory audiovisual gain as it develops across childhood.

As predicted, we found that adults benefited significantly more from visual articulation than children within this age range. Perhaps surprisingly, this could not be attributed to the inability to recognize words in the A_{alone} condition, which was subject to modest improvement with increasing age and was mostly seen at the higher SNR levels. It was apparent that even the youngest children in our sample were capable of performing this simple word recognition task. The small magnitude of performance differences in the A_{alone} condition between adults and children make it clear that differences in lexicon cannot account for the observed effects.

Children benefited substantially less from the addition of visual speech at most SNRs and this difference tended to be more pronounced as the amount of added noise was increased. Children showed a broader AV- gain function with similar levels of audiovisual gain observed across a wider range of SNRs. Our youngest sample had their gain peak at −9dBA SNR and not at −12dBA SNR, where it was found in adults and older children. Our results are in line with data published by Wightman and Kistler (2006) who showed a monotonic age effect of the release from informational masking when 6 to 16 year old children and

---

[1]We chose the median over the arithmetic mean because of its lesser susceptibility to outliers.

adults were asked to detect masked auditory targets in a sentence context. Even children between the ages of 12 to about 17 years performed somewhat poorer than adults. The procedure used was very different from the one presented here using a closed set of targets allowing higher levels of speechreading in the A_{alone} condition (about 80% in adults) and did not assess performance at SNRs that resulted in performance levels below 40% in the A_{alone} condition in adults. While it is fair to say that we find a similar overall decrease of AV enhancement (or AV-release from masking) it is not possible to compare the results of both studies regarding the dynamics of AV performance over a wide spectrum of SNRs and the development over age due to the differences in methodology.

Our data are also in line with a recent study that examined the developmental trajectory of the ability to benefit from multisensory inputs during a simple speeded reaction-time task (Brandwein et al., 2010). It is well-established that adults can respond considerably faster to paired audio-visual or audio-visual stimulation than to the unisensory constituents in isolation, and that this facilitation is more than would be predicted by simple probability summation based on the reaction time distributions to the unisensory inputs (e.g. Harrington and Peck, 1998; Molholm et al., 2002, 2006; Murray et al., 2005). In Brandwein et al. (2010), we found that a 7–9 year-old cohort showed essentially no evidence for multisensory speeding, that a 10–12 year-old cohort showed only week multisensory speeding, but that a much more adult-like pattern emerged in a 13–16 year-old group. Similarly, Barutchu and colleagues showed only very weak multisensory speeding in both a cohort of 8 year-olds and a cohort of 10 year-olds (Barutchu et al., 2010a).

We have pointed out in earlier publications (Ross et al. 2007a,b) that the largest AV-enhancement is linked to the intelligibility of the speech signal and usually occurs where about 20% of the word information can be recognized in the A_{alone} condition. We hypothesized that at this SNR critical consonant information becomes available when presented in conjunction with visual speech. This speech information is then capable of effectively disambiguating between similar words (e.g. bar, car). We suspect that this consonant information is intelligible only with the concurrent visual stimulus at −9dBA SNR rather than −12dBA SNR in the youngest children of our sample.

That increase of overall gain over age is not uniform may indicate that the development of the integration of visual and auditory speech signals proceeds in "sensitive" periods between the ages of 8 to 9 and 10 to 11. However, a closer look at the overall gain at the four lowest SNR levels, where AV-gain was the greatest, suggests a slightly more complicated picture. Multisensory gain develops quite uniformly at the lowest and the intermediate SNR levels with somewhat more variability at other SNRs. It remains unclear whether this pattern reflects genuine developmental differences as a function of SNR or whether it is simply a result of the particularities of our sample. Speculations about the underlying mechanisms should await replication of this pattern.

We mentioned that the developmental increase in AV- gain is not likely to be related to an increase in the ability to speechread. In the past, we did not find indicators of a relationship between AV- gain and speechreading except at the lowest SNR. This relationship at the lowest SNR is not surprising because this condition is similar to speechreading since at this SNR words presented in the A_{alone} condition are not intelligible unless visual articulation is added. In our previous study AV- gain in adults was not correlated with speechreading performance, in line with previous reports (e.g. Cienkowski and Carney, 2002; Gagné, Querengesser, Folkeard, Munhall, and Masterson, 1995; Munhall, 2002; Ross et al., 2007b; Watson, Qiu, Chamberlain, and Li, 1996). However, we did find that AV- gain in children was indeed correlated with V_{alone} performance, although this relationship was only present at the lowest and intermediate SNR. This is in line with findings by Massaro et al. (1986)

who suggested that speech-reading ability might have played a role in the reduced audiovisual enhancement in children. On the basis of their findings they speculated that the ability to speechread, and therefore the influence of visual information on AV speech perception, evolves over a relatively lengthy period but that it is terminated some time soon after the child's 6th year. They suggested that the beginning of schooling may be a significant factor in the increase of the influence of visual information. However, it is likely that the development of the influence of speaker articulation on speech perception increases as demands on and exposure to audiovisual speech recognition abilities increase in noisy classroom environments. Indeed, our data show that the gain from seeing speaker articulation continues to develop well into late childhood.

The ability to recognize speech in the absence of auditory input (speech-reading) and audiovisual speech perception in noise, although involving related information and therefore often erroneously equated, are not equivalent. There is considerable variability in speech-reading abilities in normal adult subjects and this ability involves a range of cognitive and perceptual abilities that are not necessarily related to speech, such as certain aspects of working memory and the ability to use semantic contextual information (Rönneberg et al., 1998). Consequently, performance on silent speech-reading does not appear to correlate with perceiving sentences in noise or McGurk-like tasks (Munhall and Vatikiotis-Bateson, 2004). Speech-reading, involving different perceptual and cognitive strategies, while challenging for children (Lyxell and Holmberg, 2000), is unlikely to be substantially related to their reduced audiovisual enhancement according to the current results.

Some authors have suggested that the increase in visual influence on AV speech perception across development is related to experience in the self-articulation of speech sounds. Desjardins et al. (1997) found that 3 to 5 year old children that made substitution errors on an articulation test also scored lower on speech-reading tests and showed less influence of visual articulation during audiovisual speech perception. Siva and colleagues found that adult patients with cerebral palsy, lacking experience with normal speech production, also showed less visual influence on AV-speech than control subjects (Siva, Stevens, Kuhl and Metzoff, 1995). However, contradictory evidence comes from a study by Dodd et al. (2008) who found that speech-disordered children and matched controls did not differ in their susceptibility to the McGurk illusion, or in their favored strategy in response to incongruent auditory and visual speech stimuli. Given the evidence from past studies, it seems reasonable to venture that the ability to derive information from the visual speech signal develops both as a function of exposure to audiovisual speech signals and as a function of the self-production of speech.

Our finding of reduced influence of visual inputs upon speech processing in children is consistent with a large extant literature (Massaro, 1984; Massaro et al., 1986; Desjardins, Rogers, and Werker, 1997; Hockley and Polka, 1994; Sekiyama and Burnham, 2008). The previous literature has also suggested that the greater part of developmental change, in terms of visual influences on auditory perception, has already occurred relatively early in childhood during the first years after entering school (see Massaro, 1984; Massaro et al. 1986). On the other hand, the current results suggest that multisensory enhancement of speech continues to increase until adolescence and perhaps into adulthood, but a definite answer to this question is reserved for future investigations. One factor for this seeming divergence with the previous literature may well pertain to differences in the nature of the tasks typically used to test these abilities. While the McGurk effect, which was used in the majority of the previous work, has proven an excellent tool for investigating multisensory speech processing and represents a compelling audiovisual illusion, it is hard to precisely determine just how the assessed visual influence translates into a realistic environmental context and it has been suggested that the McGurk effect should be distinguished from the

perception of audiovisual speech (Jordan and Sergeant, 2000). Perhaps not surprisingly, task performance in McGurk-like scenarios does not seem to be related to individual differences in audiovisual perception of sentence materials (Grant and Seitz, 1998; Munhall and Vatikiotis-Bateson, 2004)[2].

As mentioned above, perhaps the most surprising finding here was the fact that there was no change in word recognition in the auditory-alone condition between the 5–7 year-olds and 9–14 year-olds, a finding that was consistent across all signal-to-noise levels. In contrast, Eisenberg (2000) found that recognition abilities were better in a group of 10–12 year old children than in 5–7 year-olds when asked to recognize a variety of different speech material in masking noise (i.e. sentences, words, nonsense syllables and digits). Furthermore, and unlike the current findings, their 10–12 year-olds performed just as well as an adult group. The authors attributed the lower performance of younger children to an inability to fully use the sensory information available to them together with their incomplete linguistic/cognitive development. Elliott (1979) came to similar conclusions when presenting children of 9 to 17 years and adults with sentences providing high or low semantic context under varying SNRs (+5 dB, 0dB, −5dB). He found that the performance of 9-year-old children was significantly poorer than that of older children and adults in all conditions. Eleven to 13-year-old children performed significantly lower than 15–17 year olds and adults, but this decrease was confined to the high-context sentence material. When low semantic or lexical context was provided, such as in low context sentences or monosyllabic words, 11–13 year olds performed at the level of older children and adults. Elliott consequently concluded that this difference was likely due to differences in linguistic knowledge but not perceptual abilities. At younger ages (9 years and younger) the lower performance was suggested to be additionally impacted by the detrimental effects of masking noise on acoustic and perceptual processing.

Talarico et al. (2007) also reported that older children (12 to 16) outperformed younger children (6 to 8) in the identification of monosyllabic words that were masked in noise and found that cognitive abilities as assessed by the WISC-III were not related to the correct identification of the words (also Fallon et al., 2000; Nittrouer and Boothroyd, 1990). The authors suggested that age-related differences in the perception of speech-in-noise were primarily due to sensory factors. While on the face of it, the small change in the ability to detect words in noise under auditory alone conditions seems somewhat inconsistent with previous work and will certainly require replication; there is extant work that is consistent with our findings. For example, Johnson (2000) showed that the ability to identify vowels and consonants in naturally produced nonsense syllables has different developmental onsets that span the age-range used here. When syllables were embedded in multi-speaker babble, consonant identification reached adult levels at about 14 years of age. When reverberation was added to further complicate listening conditions, consonant identification did not appear to mature until the late teenage years, whereas the identification of vowels matured considerably earlier (by the age of 10). This finding has interesting implications for the results reported here. First, as we have pointed out in an earlier publication (Ross et al., 2007b), consonants are easier to mask in noise than vowels due to their lower power (Barnett, 1999; French and Steinberg, 1947). Consonants contain important information for the recognition of words since many, especially monosyllabic words, share the same vowel configuration (e.g. game, shame, blame, tame, name) and therefore have a high lexical neighborhood density. In the weak signal conditions present at lower SNRs, often only vowels are intelligible and words can therefore remain ambiguous. Although visual

---

[2]Note that differences between 11 year-olds and adults have in fact been seen using a McGurk task, suggesting that the development of audiovisual speech processes is extended into later years, but these results were only ever reported in a brief abstract (Hockley and Polka, 1994).

articulation is often redundant to the auditory signal, it often contains this critical consonant information (such as place of articulation) and it therefore serves to disambiguate fragmented information in the auditory channel. We have previously proposed that this supplemental visual information provides maximal enhancement when accompanied by a certain, critical amount of acoustic consonant information (Ross et al., 2007b).

In this study, this was the case where about 11%[3] of the words were identified in the A_{alone} condition for adults and at about 21% in (all) children. With increasing signal to noise ratio, the auditory signal then becomes increasingly intelligible on its own and we see a decreasing audiovisual enhancement. This results in the characteristic gain curve found in adults with a maximal AV- benefit at intermediate SNRs, granted that AV-enhancement is quantified as a simple difference score (AV -A). The absence of differences in A_{alone} performance between young and older children in our study may have been due to a late developing ability to identify consonants in high levels of masking noise.

It is important to note that the multisensory integration of visual and auditory consonant information, being sensory/perceptual in nature, is not the only factor explaining the differences in audiovisual integration between adults and children. The identification of words, especially in low SNRs, also clearly imposes high cognitive demands on the participant. When the auditory, audiovisual or visual percept is highly ambiguous on a perceptual level, cognitive and strategic processes operating on lexical knowledge become increasingly important (e.g. Allen and Wightman, 1992; Boothroyd, 1997; Eisenberg et al., 2000; Elliott, 1979; Elliott et al., 1979; Hnath-Chisolm, Laipply, and Boothroyd, 1998). For example, ambiguous percepts mostly require the participant to make a guess about the identity of the target word. The process of guessing relies heavily on memory functions (e.g. the rehearsal of the fragmented word), and the strategic use of lexical knowledge like the selective recall of similar sounding words from the lexicon and their probabilistic evaluation as target candidates on the basis of the given visual and auditory perceptual information. These more general cognitive factors could also be the reason for the correlation between speechreading and overall gain that we found in children. Higher cognitive/strategic abilities may have influenced speechreading and AV-performance alike. As these cognitive abilities mature with age, their relative contribution to AV-performance may diminish relative to sensory-perceptual factors, which could in turn explain why a relationship between speechreading and AV-gain was absent in adults. Support for this idea comes from recent evidence showing a relationship between general cognitive abilities (Wechsler IQ) in children and performance in a reaction time task using unisensory and multisensory stimuli (Baruthcu et al., 2010b).

Some have argued that cognitive abilities are not related to the ability to recover (unisensory) speech in noise in childhood (Fallon et al., 2000; Nittrouer and Boothroyd, 1990; Talarico et al., 2007) and that these abilities are mainly an intrinsic feature of the auditory system that matures with age. We believe, however, that cognitive factors likely played an important role in this study. Although children performed well in conditions with low or no noise, increasing noise is likely to raise cognitive demands for adults but especially for children with a highest impact at the lowest SNR. However, if cognitive factors were the main contributor to the decreased performance in children, one would expect to find an increased impact in the younger children, which was simply not the case here.

---

[3]The A_{alone} performance was 11% at −12dBA SNR and 28% at −9dBA SNR which suggest that the actual maximum AV-gain point for adults under the current experimental condition would probably have been at around −10.5 dBA SNR, a condition that was not presented in this experiment.

It should be noted that some studies have suggested a role for attention in multisensory integrative processes (Alsius et al., 2005; Barutchu et al., 2010b; Watanabe and Shimojo, 1998). It has also been shown that children are affected by noise more than adults (Elliott, 1979; Elliott, Connors, Kille, Levin, Ball, and Katz, 1979; Hetu, Truchon-Gagnon, & Bilodeau, 1990; Nittrouer and Boothroyd, 1990) and it has been argued that multisensory gain may be differentially affected in children due to their higher distractibility (Barutchu et al., 2010b). However, the current data do not accord well with a simple attentional account. If the addition of noise had simply led to a loss of attentional deployment to the task or modality in our younger cohorts, then it is unclear why we did not observe similar decrements in the auditory alone condition.

An important remaining question regards the neurophysiological correlates of the observed developmental change in the ability to benefit from visual speech. As a behavioral study, the data obtained in our experiment do not directly inform about neurobiological development. However, it is firmly established that the brain continues to mature throughout childhood (e.g., Shaw et al., 2008; Fair et al., 2008), and that these developmental changes are associated with substantial changes in cognitive function (e.g., Liston et al., 2006; Somerville and Casey, 2010). Changes are seen anatomically in the form of increases and decreases in cortical thickness (Shaw et al., 2008, Sowell et al., 2004), which have been attributed, at least in part, to increasing myelination and the process of neural pruning. Of particular note, increased functional connectivity among more distant as compared to local cortical areas has also been observed (e.g., Fair et al., 2008; Fair et al., 2009; Kelly et al., 2009; Power, Fair, Schlaggar, & Petersen, 2010). Language processing, even when it is restricted to simple word recognition, involves a highly distributed cortical network (e.g., Pulvermuller, Shtyrov, & Hauk, 2009; see Price, 2010, for a review of some of the recent imaging literature, and Pulvermuller, 2010, for discussion of the neural representation of language). Notably, these so-called perisylvian language areas show the largest developmental increase in cortical thickness, and this process continues until late childhood (Gogtay et al, 2004; Sowell, et al. 2004; O'Donnell et al., 2005; Shaw et al., 2008). A following decrease in cortical thickness of language areas in the left hemisphere has been shown to be associated with cognitive abilities, specifically verbal learning (Sowell et al., 2001), vocabulary (Sowell, 2004) and verbal fluency (Porter et al. 2011). It is reasonable to assume that these prolonged maturational changes in cortical anatomy, and extended experience with language, are associated with increases in long-range functional connectivity of these nodes, which in turn, give rise to the successful integration of visual articulatory information to boost auditory word identification in our task.

In an fMRI study of audio-visual speech perception in healthy adults, Nath and Beauchamp (2011) recently demonstrated dynamic changes in long-range functional connectivity between the superior temporal sulcus, a region that plays a role in the integration of audio-visual speech (Beauchamp, Nath, & Pasular, 2010; Calvert, 2000), and auditory and visual sensory cortices, as a function of level of noise of the auditory and visual inputs. Changes in connectivity were assumed to reflect the differential weighting of the auditory and visual speech cues as a function of their perceptual reliability (see e.g., Ernst and Banks, 2002). Not only does one expect that the development of long range connectivity influences the impact of visual inputs on audio-visual speech perception, but also that the ability to benefit from visual articulatory information during speech perception might also follow a prolonged developmental trajectory. Indeed, by comparing 8 to 10 year olds and adults, Dick and colleagues (2010) recently showed developmental changes in the functional connectivity between brain regions known to be associated with the integration of auditory and visual speech information (supramarginal gyrus) and speech-motor processing (posterior inferior frontal gyrus and the ventral premotor cortex). This development may reflect changes in the mechanisms that relate visual speech information to articulatory speech representations

through experience producing and perceiving speech. This of course would have significant implications for the development of audio-visual speech processing.

Conclusion

In this study we asked whether the characteristic tuning pattern for audiovisual enhancement of speech that has been found in adults is subject to change across development. Consistent with previous literature, we found that children experienced less multisensory enhancement provided by the visual speech signal and that younger children don't show the characteristic peak found in adolescents and adults. In our study, audiovisual gain was subject to substantial developmental change between children of the age of 5 into adulthood whereas unisensory auditory speech recognition changed little during that period. A maximal AV-enhancement at "intermediate" signal to noise ratios is absent in preschool children and during the early school years and continues to develop into adolescence.

Acknowledgments

References

Aldridge MA, Braga ES, Walton GE, Bower TG. The intermodal representation of speech in newborns. Developmental Science. 1999; 2:42–46.

Allen P, Wightman F. Spectral pattern discrimination by children. Journal of Speech and Hearing Research. 1992; 35(1):222–233. [PubMed: 1735972]

Barnett PW. Overview of speech intelligibility. Proceedings of the Institute of Acoustics. 1999; 21:1–15.

Barutchu A, Danaher J, Crewther SG, Innes-Brown H, Shivdasani MN, Paolini AG. Audiovisual integration in noise by children and adults. J Exp Child Psychol. 2010a; 105(1–2):38–50. [PubMed: 19822327]

Barutchu A, Crewther SG, Fifer J, Shivdasani MN, Innes-Brown H, Toohey S, Danaher J, Paolini AG. The relationship between multisensory integration and IQ in children. Dev Psychol. 2010b

Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. Journal of Neuroscience. 2010; 30:2414–2417. [PubMed: 20164324]

Burnham, D.; Dodd, B. Familiarity and novelty in infant cross-language studies: Factors, problems, and a possible solution. In: Lipsitt, LP.; Rovee-Collier, C.; Hayne, H., editors. Advances in infancy research. Vol. 12. Greenwood Publishing Group; 1998. p. 170-187.

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. Neural processes underlying perceptual enhancement by visual speech gestures. NeuroReport. 2003; 14:2213–2218. [PubMed: 14625450]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology. 2000; 10:649–657. [PubMed: 10837246]

Cienkowski KM, Carney AE. Auditory-visual speech perception and aging. Ear and Hearing. 2002; 23:439–449. [PubMed: 12411777]

Coltheart M. The MRC Psycholinguistic Database. Quarterly Journal of Experimental Psychology. 1981; 33A:497–505.

Desjardins RN, Rogers J, Werker JF. An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. Journal of Experimental Child Psychology. 1997; 66:85–110. [PubMed: 9226935]

Dick AS, Solodkin A, Small SL. Neural development of networks for audiovisual speech comprehension. Brain and Language. 2010; 114(2):101–104. [PubMed: 19781755]

Dodd B. Lip-reading in infants: Attention to speech presented in - and out - of synchrony. Cognitive Psychology. 1979; 11:478–484. [PubMed: 487747]

Dodd B, McIntosh B, Erdener D, Burnham D. Perception of the auditory-visual illusion in speech perception by children with phonological disorders. Clinical Linguistics and Phonetics. 2008; 22(1):69–82. [PubMed: 18092221]

Eisenberg LS, Shannon RV, Schaefer Martinez A, Wygonski J, Boothroyd A. Speech recognition with reduced spectral cues as a function of age. Journal of the Acoustical Society of America. 2000; 107(5):2704–2710. [PubMed: 10830392]

Elliott LL. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. Journal of the Acoustical Society of America. 1979; 66:651–653. [PubMed: 489836]

Elliott LL, Connors S, Kille E, Levin S, Ball K, Katz D. Children's understanding of monosyllabic nouns in quit and in noise. Journal of the Acoustical Society of America. 1979; 66:12–21. [PubMed: 489827]

Erber NP. Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech and Hearing Research. 1969; 12:423–425. [PubMed: 5808871]

Erber NP. Auditory and audiovisual reception of words in low-frequency noise by children with normal hearing and by children with impaired hearing. Journal of Speech and Hearing Research. 1971; 143:496–512. [PubMed: 5163883]

Erber NP. Auditory-visual perception in speech. Journal of Speech and Hearing Disorders. 1975; 40:481–492. [PubMed: 1234963]

Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002; 24(4):429–33. [PubMed: 11807554]

Fair DA, Cohen AL, Dosenbach NU, Church JA, Miezin FM, Barch DM, Raichle ME, Petersen SE, Schlaggar BL. The maturing architecture of the brain's default network. PNAS. 2008; 105:4028–4032. [PubMed: 18322013]

Fair DA, Cohen AL, Power JD, Dosenbach NU, Church JA, Miezin FM, Schlaggar BL, Petersen SE. Functional brain networks develop from a "local to distributed" organization. PLoS Computational Biology. 2009:5.

Fallon M, Trehub SE, Schneider BA. Children's perception of speech in multitalker babble. Journal of the Acoustical Society of America. 2000; 108(6):3023–3029. [PubMed: 11144594]

French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. Journal of the Acoustical Society of America. 1947; 19:90–119.

Gagné JP, Querengesser C, Folkeard P, Munhall K, Masterson VM. Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech. The Volta Review. 1995; 97:33–51.

Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Nugent TF 3, Herman DH, Clasen LS, Toga AW, Rapoport JL, Thompson PM. Dynamic mapping of human cortical development during childhood through early adulthood. Proc Natl Acad Sci. 2004; 10(21):8174–9. [PubMed: 15148381]

Grant KW, Seitz PF. Measures of audio-visual integration in nonsense syllables and sentences. Journal of the Acoustical Society of America. 1998; 104:2438–2450. [PubMed: 10491705]

Harrington LK, Peck CK. Spatial disparity affects visual-auditory interactions in human sensorimotor processing. Exp Brain Res. 1998; 122:247–252. [PubMed: 9776523]

Hetu R, Truchon-Gagnon C, Bilodeau S. Problems of noise in school settings: A review of the literature and the results of an explorative study. Journal of Speech- Language Pathology and Audiology. 1990; 14:31–39.

Hnath-Chisolm TE, Laipply E, Boothroyd A. Age-related changes on a children's test of sensory-level speech perception capacity. Journal of Speech, Language and Hearing Research. 1998; 41(1):94–106.

Hockley N, Polka L. A developmental study of audiovisual speech perception using the McGurk paradigm. Journal of the Acoustical Society of America. 1994; 96:3309.

Johnson CE. Children's phoneme identification in reverberation and noise. Journal of Speech, Language and Hearing Research. 2000; 43(1):144–157.

Jordan T, Sergeant P. Effects of distance on visual and audiovisual speech recognition. Language and Speech. 2000; 43:107–124.

Kelly AM, Di Martino A, Uddin LQ, Shehzad Z, Gee DG, Reiss PT, Margulies DS, Castellanos FX, Milham MP. Development of anterior cingulate functional connectivity from late childhood to early adulthood. Cerebral Cortex. 2009; 19:640–657. [PubMed: 18653667]

Kucera, H.; Francis, WN. Computational Analysis of Present-Day American English. Providence, I: Brown University Press; 1967.

Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. Science. 1982; 218:1138–1141. [PubMed: 7146899]

Kuhl PK, Meltzoff AN. The intermodal representation of speech in infants. Infant Behavior and Development. 1984; 7:361–381.

Liston C, Matalon S, Hare TA, Davidson MC, Casey BJ. Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. Neuron. 2006; 50:643–53. [PubMed: 16701213]

Lyxell B, Holmberg I. Visual speechreading and cognitive performance in hearing-impaired and normal hearing children (11–14 years). British Journal of Educational Psychology. 2000; 70(4): 505–518. [PubMed: 11191184]

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PLoS One. 2009; 4(3):e4638. [PubMed: 19259259]

MacKain K, Studdert-Kennedy M, Spieker S, Stern D. Infant intermodal speech perception is a left-hemisphere function. Science. 1983; 219:1347–1349. [PubMed: 6828865]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264 (5588):746–748. [PubMed: 1012311]

Massaro DW. Children's perception of visual and auditory speech. Child Development. 1984; 55:1777–1788. [PubMed: 6510054]

Massaro DW, Thompson LA, Barron B, Laren E. Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology. 1986; 41:93–113. [PubMed: 3950540]

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high density electrical mapping study. Cognitive Brain Res. 2002; 14:115–128.

Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ. Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. J Neurophysiol. 2006; 96:721–729. [PubMed: 16687619]

Munhall KG, Servos P, Santi A, Goodale MA. Dynamic visual speech perception in a patient with visual form agnosia. Neuroreport. 2002; 13:1793–1796. [PubMed: 12395125]

Munhall, KG.; Vatikiotis-Bateson, E. Spatial and temporal constraints on audiovisual speech perception. In: Calvert, GA.; Spence, C.; Stein, BE., editors. The Handbook of Multisensory Processes. Cambridge, MA: Bradford, MIT Press; 2004. p. 203-223.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex. 2005; 15:963–974. [PubMed: 15537674]

Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. Journal of Neuroscience. 2011; 31:1704–1714. [PubMed: 21289179]

Nittrouer S, Boothroyd A. Context effects in phoneme and word recognition by young children and older adults. Journal of the Acoustical Society of America. 1990; 87:2705–2715. [PubMed: 2373804]

O'Donnell S, Noseworthy MD, Levine B, Dennis M. Cortical thickness of the frontopolar area in typically developing children and adolescents. Neuroimage. 2005; 24(4):2929–35.

Patterson ML, Werker JF. Two-month-old infants match phonetic information in lips and voice. Developmental Science. 2003; 6:191–196.

Porter JN, Collins PF, Muetzel RL, Lim KO, Luciana M. Associations between cortical thickness and verbal fluency in childhood, adolescence and young adulthood. Neuroimage. 2011 10.1016.

Power JD, Fair DA, Schlaggar BL, Petersen SE. The development of human functional brain networks. Neuron. 2010; 67:735–748. [PubMed: 20826306]

Pulvermüller F. Brain embodiment of syntax and grammar: discrete combinatorial mechanisms spelt out in neuronal circuits. Brain Language. 2010; 112:167–179.

Pulvermüller F, Shtyrov Y, Hauk O. Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. Brain Language. 2009; 110:81–94.

Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. Annals of New York Academy of Sciences. 2010; 1191:62–88.

Rönneberg J, Andersson J, Samuelsson S, Sonderfeldt B, Lyxell B, Risberg J. A speechreading expert: The case of MM. Journal of Speech, Language and Hearing Research. 1999; 42:5–20.

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. Cerebral Cortex. 2007a; 17:1147–1153. [PubMed: 16785256]

Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ. Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophrenia Research. 2007b; 97:173–183. [PubMed: 17928202]

Saffran, JR.; Werker, J.; Werner, L. The infant's auditory world: Hearing, speech, and the beginnings of language. In: Siegler, R.; Kuhn, D., editors. Handbook of Child Development. New York: Wiley; 2006. p. 58-108.

Saint-Amour D, DeSanctis P, Molholm S, Ritter W, Foxe JJ. Seeing Voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia. 2007; 45:587–597. [PubMed: 16757004]

Sekiyama K, Burnham D. Impact of language on development of auditory-visual speech perception. Developmental Science. 2008; 11 (2):306–320. [PubMed: 18333984]

Shaw P, Kabani NJ, Lerch JP, Eckstrand K, Lenroot R, Gogtay N, Greenstein D, Clasen L, Evans A, Rapoport JL, Giedd JN, Wise SP. Neurodevelopmental Trajectories of the human cerebral cortex. The Journal of Neuroscience. 2008; 28(14):3586–3594. [PubMed: 18385317]

Siva N, Stevens E, Kuhl PK, Metzoff A. A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. Journal of the Acoustical Society of America. 1995; 98:2983.

Somerville LH, Casey BJ. Developmental neurobiology of cognitive control and motivational systems. Current Opinion in Neurobiology. 2010; 2:236–241. [PubMed: 20167473]

Sowell ER, Delis D, Stiles J, Jernigan TL. Inproved memory function and frontal lobe maturation between childhood and adolescence: a structural MRI study. J Int Neuropsychol Soc. 2001; 7(3): 312–22. [PubMed: 11311032]

Sowell ER, Peterson BS, Thompson PM, Leonard Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. Nat Neurosci. 2003; 6:309–315. [PubMed: 12548289]

Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal mapping of cortical thickness and brain growth in normal children. The Journal of Neuroscience. 2004; 22(38):8223–8231. [PubMed: 15385605]

Sumby WH, Pollack I. Visual Contribution to Speech Intelligibility in Noise. Journal of the Acoustical Society of America. 1954; 26:212–215.

Talarico M, Abdilla G, Aliferis M, Balazic I, Giaprakis I, Stefanakis T, Foenander K, Grayden DB, Paolini AG. Effect of age and cognition on childhood speech in noise perception abilities. Audiology and Neurotoxology. 2007; 12(1):13–19.

Von Economo, C.; Koskinas, GN. Die Dytoarchitektonik der Hirnrinde des erwachsenen Menschen. Berlin: Springer; 1925.

Watanabe K, Shimojo S. Attentional modulation in perception of visual motion events. Perception. 1998; 27(9):1041–1054. [PubMed: 10341934]

Walton GE, Bower TG. Amodal representations of speech in infants. Infant Behavior and Development. 1993; 16:233–243.

Watson CS, Qiu WW, Chamberlain MM, Li X. Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. Journal of the Acoustical Society of America. 1996; 100:1153–1162. [PubMed: 8759968]

Wightman F, Kistler D, Brungart D. Informational masking of speech in children: Auditory-visual integration. Journal of the Acoustical Society of Americ. 2006; 119(6):3940–3949.

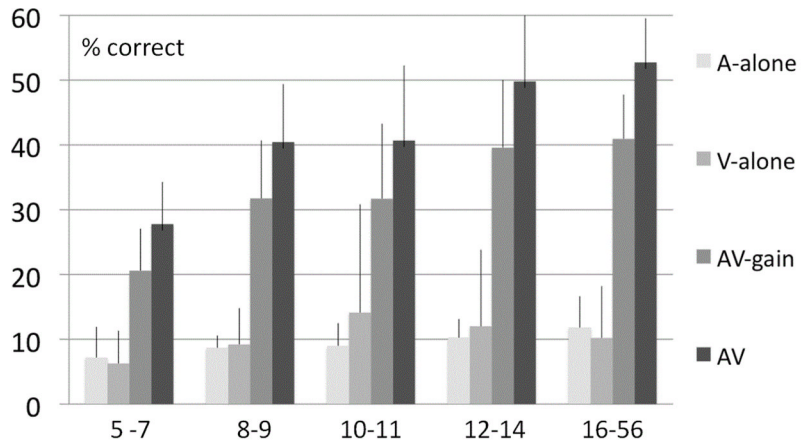**Figure 1.**
Performance in the Aalone, Valone, AV conditions and AV- gain (AV-A) averaged over the four lowest SNRs for the five age groups. Error bars depict the standard error of the mean.
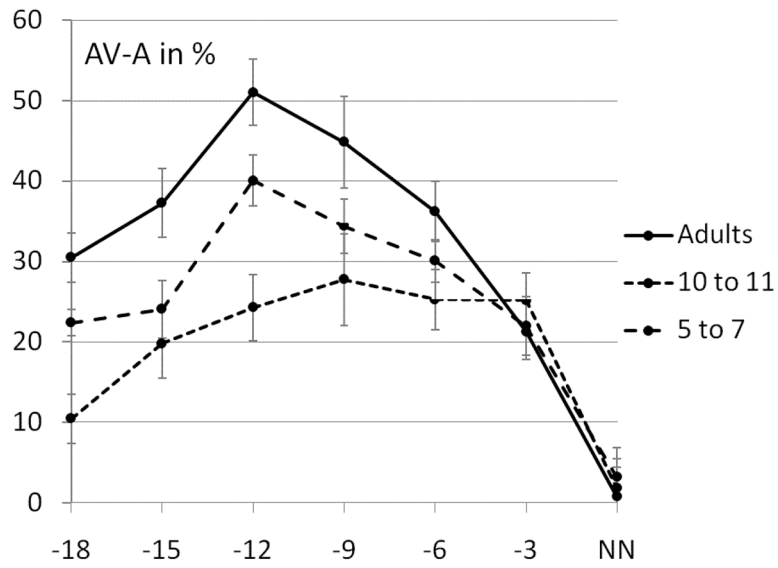
Figure 2.
Average AV- gain in % (AV-A) over all SNRs for adults and the ages 5 to 7 and 10 to 11.
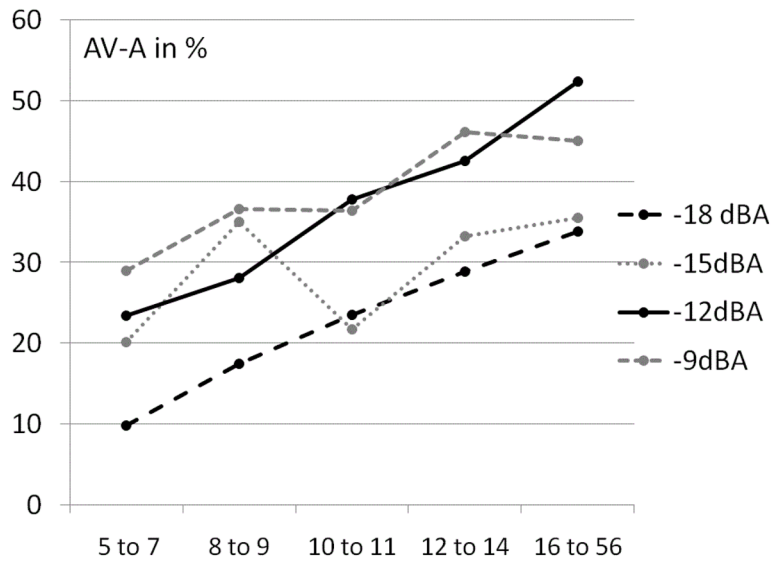Error bars represent the standard error of the mean (SE).

Figure 3.
Median AV- gain (AV-A) in % for all 5 age groups in our sample at the 4 lowest SNRs.
Monotonic linear developmental trajectories are marked in black.

Table 1

Results of the multivariate GLM. Effect of factor group on AV-gain at the four lowest SNRs.

SNR	F (1, 53)	p	η^2
−18	6.66	< 0.001*	0.33
−15	4.21	0.05*	0.24
−12	5.1	0.002*	0.28
−9	2.52	0.052	0.16

η^2: partial eta- squared.

*
$p < 0.05$.