EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, Vol. 33, pp. 2329–2337, 2011 doi:10.1111/j.1460-9568.2011.07685.x

COGNITIVE NEUROSCIENCE

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

Lars A. Ross,[1] Sophie Molholm,[1,2,3] Daniella Blanco,[1,3] Manuel Gomez-Ramirez,[4] Dave Saint-Amour[5,6] 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]Zanvyl Krieger Mind/Brain Institute Johns Hopkins University Baltimore, Maryland

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

[6]Départment de Psychologie Université du Québec à Montréal (UQAM) C.P. 8888 Succursale Centre-ville Montréal, Québec, H3c 3p8, Canada

Keywords: audiovisual, children, crossmodal, intersensory, sensory integration, speech-in-noise

Abstract

Observing a speaker's articulations substantially improves the 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 being crucial to speech and 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 (SNRs). Here, we investigated when these properties emerge during childhood by testing multisensory speech recognition abilities in typically developing children aged between 5 and 14 years, and comparing them with those of adults. By parametrically varying SNRs, we found that children benefited significantly less from observing visual articulations, displaying 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.

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 & Pollack, 1954; Erber, 1969, 1971, 1975; Ross *et al.*, 2007a,b). The magnitude of this audiovisual (AV) 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 SNRs (Sumby & Pollack, 1954; Erber, 1969, 1971, 1975; Callan *et al.*, 2003). However, recent evidence from studies investigating AV enhancement of the perception of words in healthy adults suggests that this gain tends to be largest at 'intermediate' SNRs, that is, between conditions where the auditory signal is almost perfectly audible and those where it is completely unintelligible (Ross *et al.*, 2007a,b; Ma *et al.*, 2009).

Sensitivity to coordinated AV speech inputs manifests remarkably early in development, considerably before the acquisition of language. Evidence for this has been established with an AV 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 on a given stimulus display (Dodd, 1979; Kuhl & Meltzoff, 1982, 1984; Burnham & Dodd, 1998; Patterson & Werker, 2003) or by the amplitude of sucking (Walton & Bower, 1993). Preference for congruent AV speech has been found in

Correspondence: John J. Foxe or Lars A. Ross, as above.
E-mails: john.foxe@einstein.yu.edu or lars.ross@einstein.yu.edu

Received 20 March 2011, accepted 21 March 2011

2-month-old (Patterson & Werker, 2003) and 4-month-old infants for native vowels (Kuhl & Meltzoff, 1982, 1984), and in 6-month-old children for simple syllables (MacKain *et al.*, 1983). Two-month-old infants also show a preference for a talker producing congruent over incongruent ongoing speech (Dodd, 1979; Burnham & Dodd, 1998). There is even evidence for AV matching of speech in newborns (Aldridge *et al.*, 1999). Despite the early appearance of the preference for congruent AV 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 (McGurk & MacDonald, 1976; Massaro, 1984; Massaro *et al.*, 1986; Sekiyama & 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 & MacDonald, 1976; Saint-Amour *et al.*, 2007). In their original study, McGurk & MacDonald (1976) reported that children aged 3–5 and 7–8 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 aged 4–6 and 6–10 years were less influenced by the visual articulation of an animated speaker than adults. Similarly, Hockley & 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 & Burnham (2008) showed that the AV 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 *et al.*, 1990) and are less likely to identify the last word in a sentence that is presented in multi-speaker babble (Elliott, 1979; Elliott *et al.*, 1979), which is also the case for words and sentences presented in spectral noise (Nittrouer & 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 *et al.*, 2000; Fallon *et al.*, 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 the cortical anatomy in perisylvian language areas shows a relatively long developmental trajectory, with 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 those 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, although certainly present, is clearly also weaker in infants and that it develops throughout childhood until it reaches adult levels. The aforementioned studies of AV 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 AV benefit in word recognition follows a characteristic pattern in adulthood (Ross *et al.*, 2007a,b). We have hypothesized that this pattern must emerge during childhood as acoustic, linguistic and articulatory skills continue to develop rapidly (see Saffran *et al.*, 2006 for a review) and different lexical environments are encountered. It is possible that full maturity of this system may even be delayed until adolescence. Here, we investigated whether and when these properties of AV speech integration emerge during childhood by testing multisensory speech recognition abilities in a cohort of typically developing children and adolescents across an age range of 5–14 years, and comparing their performance with that of a cohort of healthy teenagers and adults (16–46 years of age).

A set of simple predictions was made. As found by others, we expected that recognition of words presented in noise would improve with age (Elliott, 1979; Elliott *et al.*, 1979; Massaro *et al.*, 1986; Eisenberg *et al.*, 2000), 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 AV gain occurs at approximately 20% auditory-alone (A_{alone}) performance (at −12 dBA SNR in adults), and so we expected that children would also show maximal benefit when their A_{alone} performance was at about 20% word recognition. However, as this level of performance is expected to shift to lower SNrs with increasing age during childhood, it stands to reason that maximal benefit will not stabilize until A_{alone} performance nears full maturity, and we therefore 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, as 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 in establishing the age-bracket at which the emergence of a fully adult pattern would be observed.

Materials and methods

Participants

Forty-four typically developing children [age range, 5–14 years; mean, 9.59; standard deviation (SD), 2.68] and 14 neurotypical adults (age range, 16–56 years; mean, 32.36; SD, 12.43) participated in this study. Typical development was defined here according to the following criteria, as ascertained through parent interview: (i) no history of neurological, psychological or psychiatric disorders; (ii) no history of head trauma or loss of consciousness; (iii) no current or past history of psychotropic medication use; and (iv) age-appropriate grade performance. All participants were native English speakers, had normal or corrected to normal vision, and had 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 those of a sample of children of similar age revealed no differences. Informed consent was obtained from all adults, children, and their care-takers. 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 on the basis of their written-word frequency (Kucera & Francis, 1967). The subset of words used in 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 videos were digitally remastered, so that the length of the video (1.3 s) and the onset of the acoustic signal were similar for all words. Average voice onset occurred at 520 ms after video onset (SD, 30 ms). Noise onset was at the same time as the video onset, 520 ms before the beginning of the speech signal. The words were presented at approximately 50 dB(A) SPL, and six different levels of pink noise were presented simultaneously with the presentation of the words at no noise (NN), and at 53, 56, 59, 62 and 65 dB(A) SPL. In one condition, the words were presented without additional noise. The SNRs were therefore NN, and −3, −6, −9, −12, −15 and −18 dB(A) SPL. These SNRs were chosen to cover a performance range in the A_{alone} condition from 0% recognized words at the lowest SNR to almost perfect recognition performance at NN.

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

The main experiment consisted of three conditions presented in randomized order. In the A_{alone} condition, the auditory recording of the words was presented in conjunction with a still image of the speaker's face; in the AV condition, observers saw the face of the female speaker articulating the words. Finally, in the visual-alone (V_{alone}) condition, the speaker's articulations were seen 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 was 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 ensured that eye fixation was maintained by reminding participants, if necessary. If a word was not clearly understood, the participant was asked to guess which word was presented. The experimenter was seated at approximately 1 m 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 (5–7 years, $n = 10$; 8–9 years, $n = 11$; 10–11 years, $n = 13$; 12–14 years, $n = 10$; 16–56 years, $n = 14$), and subjected percentage correct responses to a repeated measures ANOVA with factors of stimulus condition (A_{alone} vs. AV) and SNR level (seven levels), and the between-subjects factor of age group (five 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,b) 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 determine whether age differences in AV gain were manifested differently across SNR levels, we conducted a multivariate ANOVA 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.

AV enhancement (or AV gain) was operationalized here as the difference in performance between the AV and the A_{alone} conditions ($AV - A_{alone}$). The issue of how to characterize AV benefit is somewhat contentious and has been discussed more exhaustively elsewhere (see Holmes, 2007; Ross *et al.*, 2007a,b; Sumby & Pollack, 1954). We therefore limit ourselves here to a brief explanation of our reason for using simple difference scores as an index of AV benefit.

Different methods for characterizing AV 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 the seminal paper by Sumby & 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 it stands in stark contrast to what we know about the physiological underpinnings of AV integration (Stein and Meredith, 1993). Use of the simple difference score has the advantage that such artefacts are avoided and that it can therefore be used without concern at low SNRs, where improvement is not constrained by a performance ceiling. Although we assessed auditory, visual 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 gain 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 gain is related to speechreading ability (Massaro, 1984; Massaro *et al.*, 1986), although it bears mentioning that others have shown no correlation (e.g. Gagné *et al.*, 1995; Watson *et al.*, 1996; Cienkowski & Carney, 2002; Munhall *et al.*, 2002; Ross *et al.*, 2007b).

Results

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

The addition of visual articulation reliably enhanced word recognition performance (Sumby & 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 & 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 gain was dependent on the level of noise.

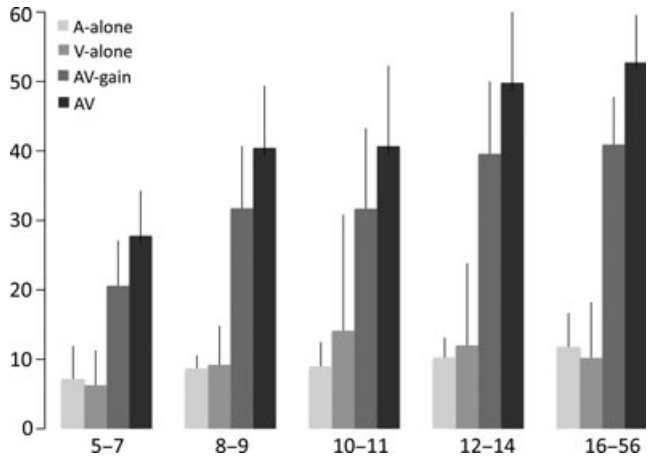

FIG. 1. Performance in the A_{alone}, V_{alone} and AV conditions and AV gain averaged over the four lowest SNRs for the five age groups. Error bars depict the standard error of the mean.

The five groups differed significantly in overall performance ($F_{4,53} = 14.62$, $P < 0.001$, $\eta^2 = 0.53$). Figure 1 shows 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 years) to 52.7% in adults.

The ability to recognize words in the A_{alone} condition changed surprisingly little across age groups, with average levels of 7.1% in the 5–7-year group 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 use of a multivariate ANOVA revealed that group differences were only apparent at higher SNR levels (−9 dB(A), $F_{4,52} = 3.39$, $P = 0.02$; −6 dB(A), $F_{4,52} = 3.08$, $P = 0.02$; −3 dB(A), $F_{4,52} = 4.71$, $P < 0.01$; 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 that is apparent in Fig. 1. Although there was a substantial increase in AV performance from the 5–7-year group to the 8–9-year group, there was very little difference in AV gain from the 8–9-year group to the 10–11-year group. In contrast, there was a substantial increase in AV gain in the 12–14-year group, and this approached adult levels of AV gain.

This observed increase in AV gain with age did not depend on an increase in the ability to speechread (V_{alone}). Speechreading performance did not increase with age, as the bar graph in Fig. 1 clearly shows. 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 result from overall performance differences among individual children, because V_{alone} performance did not covary with performance in the A_{alone} (collapsed over the four lowest SNRs) condition ($r = 0.056$, $P = 0.725$).

AV gain and SNR

We replicated the findings originally demonstrated by Ross *et al.* (2007a), showing that the magnitude of AV gain in word recognition was critically dependent on SNR, showing maximal enhancement at 'intermediate' SNR levels in adults (−12 dBA SNR) (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 (Fig. 2); however, the enhancement at each individual SNR was considerably lower than in adults. In the 5–7-year group, AV gain also increased with increasing SNR, but the overall gain was lower and the characteristic peak at −12 dBA SNR was missing, an overall shallower gain curve with a maximum at −9 dBA SNR being seen.

Developmental change in AV gain at different SNRs

We further explored whether AV performance developed uniformly at all SNR levels. We plotted AV gain (median; we chose the median over the arithmetic mean because of its lesser susceptibility to outliers) for the four lowest SNRs across the five age groups (Fig. 3). AV gain at all four SNR levels increased linearly between the youngest age group and adults. This increase was particularly stable at −12 dB(A) SNR and the lowest SNR at –18 dB(A).

This pattern was confirmed by use of a multivariate GLM, which showed that AV gain was consistent over all age groups and was manifest at all of the four lowest SNRs (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 into adulthood. We expected that

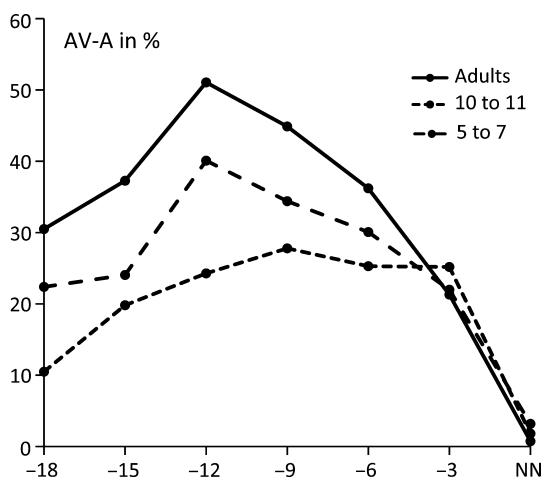

FIG. 2. Average AV gain (AV-A) over all SNRs for adults and the ages 5–7 and 10–11 years. Error bars represent the standard error of the mean.

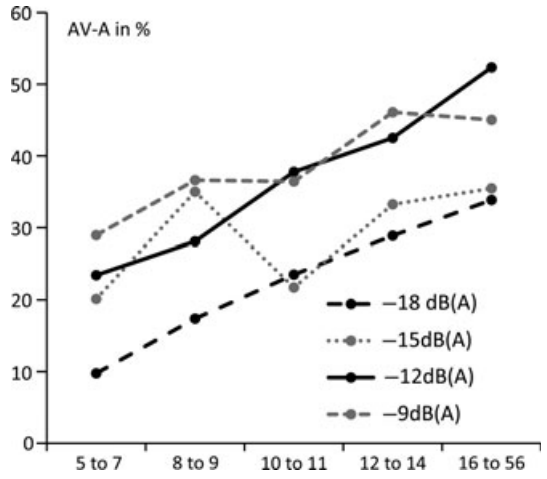

FIG. 3. Median AV gain (AV-A) for all five age groups in our sample at the four lowest SNRs.

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; *$P < 0.05$.

younger children would have more difficulty in perceiving speech in masking noise, and that, whereas they would certainly benefit from exposure to visualized speaker articulations, this benefit would be significantly lower than that seen in adults. Our primary interest was to quantify the enhancement of multisensory AV 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 makes 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 AV gain being observed across a wider range of SNRs. Our youngest sample had their gain peak at −9 dB(A) SNR and not at −12 dB(A) SNR, where it was found in adults and older children. Our results are in line with data published by Wightman *et al.* (2006), who showed a monotonic age effect of the release from informational masking when 6–16-year-old children and adults were asked to detect masked auditory targets in a sentence context. Even children between the ages of 12 and about 17 years performed somewhat more poorly 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. Although it is fair to say that we found a similar overall decrease of AV gain (or AV release from masking), it is not possible to compare the results of the two studies regarding the dynamics of AV performance over a wide spectrum of SNRs and the development over age, owing 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 AV or AV 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 & 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 of 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 *et al.* (2010a) showed only very weak multisensory speeding in both a cohort of 8-year-olds and a cohort of 10-year-olds.

We have pointed out in earlier publications (Ross *et al.*, 2007a,b) that the largest AV gain is linked to the intelligibility of the speech signal, and usually occurs when 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 then makes it possible to effectively disambiguate between similar words (e.g. bar and car). We suspect that this consonant information is intelligible only with the concurrent visual stimulus at −9 dB(A) SNR, and not at −12 dB(A) SNR, in the youngest children of our sample.

The fact that the increase of overall gain with 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 8–9 and 10–11 years. 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, and shows 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, as this condition is similar to speechreading, because, 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. Gagné *et al.*, 1995; Watson *et al.*, 1996; Cienkowski & Carney, 2002; Munhall *et al.*, 2002; Ross *et al.*, 2007b). 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 SNRs. This is in line with findings by Massaro *et al.* (1986), who suggested that speechreading ability might have played a role in the reduced AV gain 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 sixth 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 AV 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 abilities to recognize speech in the absence of auditory input (speechreading) and AV speech perception in noise, although they involve related information and are therefore often erroneously equated, are not equivalent. There is considerable variability in speechreading ability 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.*, 1999). Consequently, performance on silent speechreading does not appear to correlate with perceiving sentences in noise or McGurk-like tasks (Munhall & Vatikiotis-Bateson, 2004). Speechreading, involving different perceptual and cognitive strategies, although challenging for children (Lyxell & Holmberg, 2000), is unlikely to be substantially related to their reduced AV enhancement, according to the current results.

Some authors have suggested that the increase in visual influence on AV speech perception throughout development is related to experience in the self-articulation of speech sounds. Desjardins *et al.* (1997) found that 3–5-year-old children who made substitution errors on an articulation test also scored lower on speechreading tests and showed less influence of visual articulation during AV speech perception. Siva *et al.* 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 *et al.*, 1995). However, contradictory evidence came 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 suggest that the ability to derive information from the visual speech signal develops both as a function of exposure to AV speech signals and as a function of the self-production of speech.

Our finding of reduced influence of visual inputs on speech processing in children is consistent with a large extant literature (Massaro, 1984; Massaro *et al.*, 1986; Hockley & Polka, 1994; Desjardins *et al.*, 1997; Sekiyama & 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 reason for this seeming divergence from the previous literature may well pertain to differences in the nature of the tasks typically used to test these abilities. Although the McGurk effect, which was used in the majority of the previous work, has proven to be an excellent tool for investigating multisensory speech processing, and represents a compelling AV 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 AV speech (Jordan & Sergeant, 2000). Perhaps not surprisingly, task performance in McGurk-like scenarios does not seem to be related to individual differences in AV perception of sentence materials (Grant & Seitz, 1998; Munhall & Vatikiotis-Bateson, 2004). [Note that differences between 11-year-olds and adults have, in fact, been seen with the use of a McGurk task, suggesting that the development of AV speech processes is extended into later years, but these results were only ever reported in a brief abstract (Hockley & Polka, 1994).]

As mentioned above, perhaps the most surprising finding here was the fact that there was no change in word recognition in the A_{alone} condition between the 5–7-year-olds and the 9–14-year-olds, a finding that was consistent across all SNR levels. In contrast, Eisenberg *et al.* (2000) found that recognition abilities were better in a group of 10–12-year-old children than in 5–7-year-olds when they were asked to recognize a variety of different speech materials 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 9–17-year-olds and adults with sentences providing high or low semantic context under varying SNRs (+5, 0, and −5 dBA). He found that the performance of 9-year-olds was significantly poorer than that of older children and adults in all conditions. Children aged 11–13 years performed significantly more poorly 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 probably attributabl to differences in linguistic knowledge but not perceptual abilities. At younger ages (9 years and younger), the lower performance was suggested to be additionally affected by the detrimental effects of masking noise on acoustic and perceptual processing.

Talarico *et al.* (2007) also reported that older children (12–16 years) outperformed younger children (6–8 years) 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 Nittrouer & Boothroyd, 1990; Fallon *et al.*, 2000). The authors suggested that age-related differences in the perception of speech-in-noise were primarily attributable to sensory factors. Although, on the face of it, the small change in the ability to detect words in noise under A_{alone} conditions seems to be 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 the 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 years). 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 are vowels, owing to their lower power (French & Steinberg, 1947; Barnett, 1999). Consonants contain important information for the recognition of words, as many, especially monosyllabic, words share the same vowel configuration (e.g. game, shame, blame, tame, and name) and therefore have a high lexical neighborhood density. In the weak signal conditions present at lower SNRs, it is often

the case that only vowels are intelligible, and words can therefore remain ambiguous. Although visual 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% of the words were identified in the A_{alone} condition for adults and at about 21% in (all) children. (The A_{alone} performance was 11% at −12 dB(A) SNR and 28% at −9 dB(A) SNR, which suggests that the actual maximum AV gain point for adults under the current experimental condition would probably have been at around −10.5 dB(A) SNR, a condition that was not presented in this experiment.) With increasing SNR, the auditory signal then becomes increasingly intelligible on its own, and we see a decrease in AV gain. This results in the characteristic gain curve found in adults, with a maximal AV benefit at intermediate SNRs, granted that AV gain is quantified as a simple difference score ($AV - A_{alone}$). The absence of differences in A_{alone} performance between young and older children in our study may have resulted from 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 AV 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, AV or visual percept is highly ambiguous on a perceptual level, cognitive and strategic processes operating on lexical knowledge become increasingly important (e.g. Allen & Wightman, 1992; Boothroyd, 1997; Eisenberg *et al.*, 2000; Elliott, 1979; Elliott *et al.*, 1979; Hnath-Chisolm *et al.*, 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, such as 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 (Barutchu *et al.*, 2010b).

Some have argued that cognitive abilities are not related to the ability to recover (unisensory) speech-in-noise in childhood (Nittrouer & Boothroyd, 1990; Fallon *et al.*, 2000; Talarico *et al.*, 2007), and that these abilities mainly constitute an intrinsic feature of the auditory system that matures with age. We believe, however, that cognitive factors probably played an important role in this study. Although children performed well in conditions with low or no noise, increasing that noise is likely to raise cognitive demands not only for adults, but especially for children, with a highest impact at the lowest SNR. However, if cognitive factors constituted 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.

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 & Shimojo, 1998). It has also been shown that children are affected more than adults by noise (Elliott, 1979; Elliott *et al.*, 1979; Hetu *et al.*, 1990; Nittrouer & Boothroyd, 1990), and it has been argued that multisensory gain may be differentially affected in children, because of 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 A_{alone} condition.

An important remaining question regards the neurophysiological correlates of the observed developmental change in the ability to benefit from visual speech. As our experiment was a behavioral study, the data obtained do not provide direct information about neurobiological development. However, it is firmly established that the brain continues to mature throughout childhood (e.g., Fair *et al.*, 2008; Shaw *et al.*, 2008), and that these developmental changes are associated with substantial changes in cognitive function (e.g., Liston *et al.*, 2006; Somerville & Casey, 2010). Changes are seen anatomically in the form of increases and decreases in cortical thickness (Sowell *et al.*, 2004; Shaw *et al.*, 2008), which have been attributed, at least in part, to increasing myelination and the process of neural pruning. It is of particular note that increased functional connectivity among more distant, as compared with local, cortical areas has also been observed (e.g., Fair *et al.*, 2008, 2009; Kelly *et al.*, 2009; Power *et al.*, 2010). Language processing, even when it is restricted to simple word recognition, involves a highly distributed cortical network (e.g., Pulvermüller *et al.*, 2009; see Price, 2010; for a review of some of the recent imaging literature, and Pulvermüller, 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 subsequent 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 *et al.*, 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 a functional magnetic resonance imaging study of AV speech perception in healthy adults, Nath & 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 AV speech (Calvert *et al.*, 2000; Beauchamp *et al.*, 2010), 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 & Banks, 2002). One would expect not only that the development of long-range connectivity would influence the impact of visual inputs on AV 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 8–10-year-olds and adults, Dick *et al.* (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 ventral premotor cortex). This development may reflect changes in the mechanisms that relate visual speech information to articulatory speech representations through experience in producing and perceiving speech. This, of course, would have significant implications for the development of AV speech processing.

Conclusion

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

Acknowledgements

This study was supported by a grant from the US National Institute of Mental Health (NIMH) to J. J. Foxe and S. Molholm (RO1 – MH085322), and by pilot grants from Cure Autism Now (to J. J. Foxe) and The Wallace Research Foundation (to J. J Foxe and S. Molholm). Support to Dave Saint-Amour was provided by the Canadian Psychiatric Research Foundation.

Abbreviations

$A_{alone,}$ auditory-alone; AV, audiovisual; GLM, general linear model; NN, no noise; SD, standard deviation; SNR, signal-to-noise ratio; $V_{alone,}$ visual-alone.

References

Aldridge, M.A., Braga, E.S., Walton, G.E. & Bower, T.G. (1999) The intermodal representation of speech in newborns. *Dev. Sci.*, **2**, 42–46.

Allen, P. & Wightman, F. (1992) Spectral pattern discrimination by children. *J. Speech Hear. Res.*, **35**, 222–233.

Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco (2005). Audiovisual integration of speech falters under high attentional demands. *Curr. Biol.*, **10**, 839–843.

Barnett, P.W. (1999) Overview of speech intelligibility. *Proc. Inst. Acoust.*, **21**, 1–15.

Barutchu, A., Danaher, J., Crewther, S.G., Innes-Brown, H., Shivdasani, M.N. & Paolini, A.G. (2010a) Audiovisual integration in noise by children and adults. *J. Exp. Child Psychol.*, **105**, 38–50.

Barutchu, A., Crewther, S.G., Fifer, J., Shivdasani, M.N., Innes-Brown, H., Toohey, S., Danaher, J. & Paolini, A.G. (2010b) The relationship between multisensory integration and IQ in children. *Dev. Psychol.*, doi: 10.1037/a0021903 [Epub ahead of print].

Brandwein, A.B., Foxe, J.J., Russo, N.N., Altschuler, T.S., Gomes, H. & Molholm, S. (2010). The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. *Cereb. Cortex*, doi: 10.1093/cercor/bhq170 [Epub ahead of print].

Beauchamp, M.S., Nath, A.R. & Pasalar, S. (2010) fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J. Neurosci.*, **30**, 2414–2417.

Boothroyd, A. (1997) Auditory development of the hearing child. *Scand. Audiol. Suppl.*, **46**, 9–16.

Burnham, D. & Dodd, B. (1998) Familiarity and novelty in infant cross-language studies: factors, problems, and a possible solution. In Lipsitt, L.P., Rovee-Collier, C. & Hayne, H. (Eds), *Advances in Infancy Research, Vol. 12*. Greenwood Publishing Group, Stamford, CT: Ablex, pp. 170–187.

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**, 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. *Curr. Biol.*, **10**, 649–657.

Cienkowski, K.M. & Carney, A.E. (2002) Auditory–visual speech perception and aging. *Ear Hear.*, **23**, 439–449.

Coltheart, M. (1981) The MRC psycholinguistic database. *Q. J. Exp. Psychol.*, **33A**, 497–505.

Desjardins, R.N., Rogers, J. & Werker, J.F. (1997) An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *J. Exp. Child Psychol.*, **66**, 85–110.

Dick, A.S., Solodkin, A. & Small, S.L. (2010) Neural development of networks for audiovisual speech comprehension. *Brain Lang.*, **114**, 101–104.

Dodd, B. (1979) Lip-reading in infants: attention to speech presented in - and out - of synchrony. *Cogn. Psychol.*, **11**, 478–484.

Dodd, B., McIntosh, B., Erdener, D. & Burnham, D. (2008) Perception of the auditory–visual illusion in speech perception by children with phonological disorders. *Clin. Linguist. Phon.*, **22**, 69–82.

Eisenberg, L.S., Shannon, R.V., Schaefer Martinez, A., Wygonski, J. & Boothroyd, A. (2000) Speech recognition with reduced spectral cues as a function of age. *J. Acoust. Soc. Am.*, **107**, 2704–2710.

Elliott, L.L. (1979) Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *J. Acoust. Soc. Am.*, **66**, 651–653.

Elliott, L.L., Connors, S., Kille, E., Levin, S., Ball, K. & Katz, D. (1979) Children's understanding of monosyllabic nouns in quite and in noise. *J. Acoust. Soc. Am.*, **66**, 12–21.

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

Erber, N.P. (1971) Auditory and audiovisual reception of words in low-frequency noise by children with normal hearing and by children with impaired hearing. *J. Speech Hear. Res.*, **143**, 496–512.

Erber, N.P. (1975) Auditory–visual perception in speech. *J. Speech Hear. Disord.*, **40**, 481–492.

Ernst, M.O. & Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **24**, 429–433.

Fair, D.A., Cohen, A.L., Dosenbach, N.U., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E. & Schlaggar, B.L. (2008) The maturing architecture of the brain's default network. *Proc. Natl Acad. Sci. USA*, **105**, 4028–4032.

Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U., Church, J.A., Miezin, F.M., Schlaggar, B.L. & Petersen, S.E. (2009) Functional brain networks develop from a 'local to distributed' organization. *PLoS Comput. Biol.*, **5**, e1000381.

Fallon, M., Trehub, S.E. & Schneider, B.A. (2000) Children's perception of speech in multitalker babble. *J. Acoust. Soc. Am.*, **108**, 3023–3029.

French, N.R. & Steinberg, J.C. (1947) Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.*, **19**, 90–119.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K. & Masterson, V.M. (1995) Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: an investigation of conversational and clear speech. *Volta Rev.*, **97**, 33–51.

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F. 3rd, Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L. & Thompson, P.M. (2004) Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl Acad. Sci. USA*, **10**, 8174–8179.

Grant, K.W. & Seitz, P.F. (1998) Measures of audio-visual integration in nonsense syllables and sentences. *J. Acoust. Soc. Am.*, **104**, 2438–2450.

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

Hetu, R., Truchon-Gagnon, C. & Bilodeau, S. (1990) Problems of noise in school settings: a review of the literature and the results of an explorative study. *J. Speech Lang. Pathol. Audiol.*, **14**, 31–39.

Hnath-Chisolm, T.E., Laipply, E. & Boothroyd, A. (1998) Age-related changes on a children's test of sensory-level speech perception capacity. *J. Speech Lang. Hear. Res.*, **41**, 94–106.

Hockley, N. & Polka, L. (1994) A developmental study of audiovisual speech perception using the McGurk paradigm. *J. Acoust. Soc. Am.*, **96**, 3309.

Holmes, N.P. (2007) The law of inverse effectiveness in neurons and behavior: multisensory integration versus normal variability. *Neuropsychologia*, **45**, 3340–3345.
Johnson, C.E. (2000) Children's phoneme identification in reverberation and noise. *J. Speech Lang. Hear. Res.*, **43**, 144–157.
Jordan, T. & Sergeant, P. (2000) Effects of distance on visual and audiovisual speech recognition. *Lang. Speech*, **43**, 107–124.
Kelly, A.M., Di Martino, A., Uddin, L.Q., Shehzad, Z., Gee, D.G., Reiss, P.T., Margulies, D.S., Castellanos, F.X. & Milham, M.P. (2009) Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cereb. Cortex*, **19**, 640–657.
Kucera, H. & Francis, W.N. (1967) *Computational Analysis of Present-day American English*. Brown University Press, Providence, RI.
Kuhl, P.K. & Meltzoff, A.N. (1982) The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.
Kuhl, P.K. & Meltzoff, A.N. (1984) The intermodal representation of speech in infants. *Infant Behav. Dev.*, **7**, 361–381.
Liston, C., Matalon, S., Hare, T.A., Davidson, M.C. & Casey, B.J. (2006) Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. *Neuron*, **50**, 643–653.
Lyxell, B. & Holmberg, I. (2000) Visual speechreading and cognitive performance in hearing-impaired and normal hearing children (11–14 years). *Br. J. Educ. Psychol.*, **70**, 505–518.
Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J. & Parra, L.C. (2009) Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS ONE*, **4**, e4638.
MacKain, K., Studdert-Kennedy, M., Spieker, S. & Stern, D. (1983) Infant intermodal speech perception is a left-hemisphere function. *Science*, **219**, 1347–1349.
Massaro, D.W. (1984) Children's perception of visual and auditory speech. *Child Dev.*, **55**, 1777–1788.
Massaro, D.W., Thompson, L.A., Barron, B. & Laren, E. (1986) Developmental changes in visual and auditory contributions to speech perception. *J. Exp. Child Psychol.*, **41**, 93–113.
McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, 746–748.
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. *Cogn. Brain Res.*, **14**, 115–128.
Molholm, S., Sehatpour, P., Mehta, A.D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J.P., Schwartz, T.H. & Foxe, J.J. (2006) Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *J. Neurophysiol.*, **96**, 721–729.
Munhall, K.G. & Vatikiotis-Bateson, E. (2004) Spatial and temporal constraints on audiovisual speech perception. In Calvert, G.A., Spence, C. & Stein, B.E. (Eds), *The Handbook of Multisensory Processes*. Bradford, MIT Press, Cambridge, MA, pp. 177–188.
Munhall, K.G., Servos, P., Santi, A. & Goodale, M.A. (2002) Dynamic visual speech perception in a patient with visual form agnosia. *Neuroreport*, **13**, 1793–1796.
Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2005) Grabbing your ear: rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex*, **15**, 963–974.
Nath, A.R. & Beauchamp, M.S. (2011) Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.*, **31**, 1704–1714.
Nittrouer, S. & Boothroyd, A. (1990) Context effects in phoneme and word recognition by young children and older adults. *J. Acoust. Soc. Am.*, **87**, 2705–2715.
O'Donnell, S., Noseworthy, M.D., Levine, B. & Dennis, M. (2005) Cortical thickness of the frontopolar area in typically developing children and adolescents. *Neuroimage*, **24**, 2929–2935.
Patterson, M.L. & Werker, J.F. (2003) Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.*, **6**, 191–196.
Porter, J.N., Collins, P.F., Muetzel, R.L., Lim, K.O. & Luciana, M. (2011) Associations between cortical thickness and verbal fluency in childhood, adolescence and young adulthood. *Neuroimage*, **55**, 1865–1877.
Power, J.D., Fair, D.A., Schlaggar, B.L. & Petersen, S.E. (2010) The development of human functional brain networks. *Neuron*, **67**, 735–748.
Price, C.J. (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. NY Acad. Sci.*, **1191**, 62–88.
Pulvermüller, F. (2010) Brain embodiment of syntax and grammar: discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain Lang.*, **112**, 167–179.
Pulvermüller, F., Shtyrov, Y. & Hauk, O. (2009) Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang.*, **110**, 81–94.
Rönneberg, J., Andersson, J., Samuelsson, S., Sonderfeldt, B., Lyxell, B. & Risberg, J. (1999) A speechreading expert: the case of MM. *J. Speech Lang. Hear. Res.*, **42**, 5–20.
Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C. & Foxe, J.J. (2007a) Do you see what i am saying? Exploring visual enhancement of speech comprehension in noisy environments *Cereb. Cortex*, **17**, 1147–1153.
Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C. & Foxe, J.J. (2007b) Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.*, **97**, 173–183.
Saffran, J.R., Werker, J. & Werner, L. (2006) The infant's auditory world: hearing, speech, and the beginnings of language. In Siegler, R. & Kuhn, D. (Eds), *Handbook of Child Development*. Wiley, New York, pp. 58–108.
Saint-Amour, D., DeSanctis, P., Molholm, S., Ritter, W. & Foxe, J.J. (2007) Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, **45**, 587–597.
Sekiyama, K. & Burnham, D. (2008) Impact of language on development of auditory–visual speech perception. *Dev. Sci.*, **11**, 306–320.
Shaw, P., Kabani, N.J., Lerch, J.P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J.L., Giedd, J.N. & Wise, S.P. (2008) Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.*, **28**, 3586–3594.
Siva, N., Stevens, E., Kuhl, P.K. & Metzoff, A. (1995) A comparison between cerebral-palsied and normal adults in the perception of auditory–visual illusions. *J. Acoust. Soc. Am.*, **98**, 2983.
Somerville, L.H. & Casey, B.J. (2010) Developmental neurobiology of cognitive control and motivational systems. *Curr. Opin. Neurobiol.*, **2**, 236–241.
Sowell, E.R., Delis, D., Stiles, J. & Jernigan, T.L. (2001) Inproved memory function and frontal lobe maturation between childhood and adolescence: a structural MRI study. *J. Int. Neuropsychol. Soc.*, **7**, 312–322.
Sowell, E.R., Peterson, B.S., Thompson, P.M., Leonard, C.M., Welcome, S.E., Henkenius, A.L. & Toga, A.W. (2004) Mapping cortical change across the human life span. *Nat. Neurosci.*, **6**, 309–315.
Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E. & Toga, A.W. (2004) Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.*, **22**, 8223–8231.
Stein, B.E. & Meredith, M.A. (1993) *The Merging of the Senses*. MIT Press, Cambridge, MA.
Sumby, W.H. & Pollack, I. (1954) Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.*, **26**, 212–215.
Talarico, M., Abdilla, G., Aliferis, M., Balazic, I., Giaprakis, I., Stefanakis, T., Foenander, K., Grayden, D.B. & Paolini, A.G. (2007) Effect of age and cognition on childhood speech in noise perception abilities. *Audiol. Neurotoxol.*, **12**, 13–19.
Walton, G.E. & Bower, T.G. (1993) Amodal representations of speech in infants. *Infant Behav. Dev.*, **16**, 233–243.
Watanabe, K. & Shimojo, S. (1998) Attentional modulation in perception of visual motion events. *Perception*, **27**, 1041–1054.
Watson, C.S., Qiu, W.W., Chamberlain, M.M. & Li, X. (1996) Auditory and visual speech perception: confirmation of a modality-independent source of individual differences in speech recognition. *J. Acoust. Soc. Am.*, **100**, 1153–1162.
Wightman, F., Kistler, D. & Brungart, D. (2006) Informational masking of speech in children: auditory–visual integration. *J. Acoust. Soc. Am.*, **119**, 3940–3949.