JSLHR

Research Note

Audiovisual Integration in Children Listening to Spectrally Degraded Speech

David W. Maidment,[a] Hi Jee Kang,[a] Hannah J. Stewart,[a] and Sygal Amitay[a]

Purpose: The study explored whether visual information improves speech identification in typically developing children with normal hearing when the auditory signal is spectrally degraded.
Method: Children (n = 69) and adults (n = 15) were presented with noise-vocoded sentences from the Children's Co-ordinate Response Measure (Rosen, 2011) in auditory-only or audiovisual conditions. The number of bands was adaptively varied to modulate the degradation of the auditory signal, with the number of bands required for approximately 79% correct identification calculated as the threshold.
Results: The youngest children (4- to 5-year-olds) did not benefit from accompanying visual information, in comparison to 6- to 11-year-old children and adults. Audiovisual gain also increased with age in the child sample.
Conclusions: The current data suggest that children younger than 6 years of age do not fully utilize visual speech cues to enhance speech perception when the auditory signal is degraded. This evidence not only has implications for understanding the development of speech perception skills in children with normal hearing but may also inform the development of new treatment and intervention strategies that aim to remediate speech perception difficulties in pediatric cochlear implant users.

It is well established that speech perception is a multisensory experience. When auditory and visual sources of verbal information are presented simultaneously, visual cues often influence the perception and comprehension of stimuli in the auditory modality (Sumby & Pollack, 1954). For example, when the auditory and visual information do not match, auditory stimuli can be misperceived, as in the well-documented McGurk effect (McGurk & MacDonald, 1976): An auditory signal /ba/ simultaneously presented with a visual /ga/ often results in an illusionary percept /da/. Furthermore, a large body of empirical work has shown that viewing a speaker's mouth movements provides additional information that can improve auditory speech perception, particularly when the auditory signal is masked by background sounds (e.g., Bishop & Miller, 2009; McGettigan et al., 2012; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sánchez-García, Alsius, Enns, & Soto-Faraco, 2011; Summerfield, MacLeod, McGrath, & Brooke, 1989).

Degradation of the speech signal itself, as experienced by listeners with hearing impairment, represents another type of adverse listening that can be potentially enhanced by visual information. Prosthetic devices such as hearing aids and cochlear implants aim to improve audibility, but the auditory speech signal transmitted to the user is still degraded and is therefore suboptimal for accurate speech perception. Visual speech cues therefore provide one of the main ways in which individuals who are deaf or hearing impaired can access the spoken signal (Bernstein & Grant, 2009; Grant, Walden, & Seitz, 1998; Lachs, Pisoni, & Kirk, 2001; Walden, Prosek, & Worthington, 1974, 1975). Indeed, even many years postimplantation, a bias appears to exist for users of cochlear implants toward a speaker's mouth movements when integrating incongruent auditory and visual speech (as evidenced by an increased McGurk effect) in comparison to individuals with normal hearing (Rouger, Fraysse, Deguine, & Barone, 2008; Rouger et al., 2007; Schorr, Fox, van Wassenhove, & Knudsen, 2005). Presumably, one of the reasons that cochlear implant users maintain a dependence on a speaker's mouth movements following implantation is because visual cues provided by speechreading continue to provide a reliable source of nonredundant verbal information.

Studies examining the developmental time course of audiovisual integration have shown that children are sensitive to visual speech cues from a very young age, even before language acquisition. In fact, visual information has been shown to be important between 4 and 8 months of

[a]MRC Institute of Hearing Research, Nottingham, United Kingdom
Correspondence to David W. Maidment: dmaidment@ihr.mrc.ac.uk

Editor: Jody Kreiman
Associate Editor: Ewa Jacewicz
Received February 12, 2014
Revision received May 22, 2014
Accepted September 3, 2014
DOI: 10.1044/2014_JSLHR-S-14-0044

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

age—a time in development when associations between auditory and visual speech signals are advantageous to language learning (Lewkowicz & Hansen-Tift, 2012). Furthermore, using habituation and dishabituation paradigms, McGurk effects have been observed in infants as young as 4 months old (Burnham & Dodd, 1996, 2004; Rosenblum, Schmuckler, & Johnson, 1997). However, although infants appear to be able to identify whether information across spoken auditory and visual channels does or does not match, the ability to combine both inputs to derive a benefit and improve perception appears to develop much later (Barutchu et al., 2011). Although speech perception is improved in children from 4 years of age when visual information is also present in a quiet environment (Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986), the magnitude of gain is smaller than that seen in adults (Desjardins & Werker, 2004; Hockley & Polka, 1994; Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008). Similarly, typically developing children up to 14 years of age benefit less than adults from observing visual articulations when speech sounds are presented in noise (e.g., Barutchu et al., 2010; Ross et al., 2011; Wightman, Kistler, & Brungart, 2006), but the benefit grows with age and continues to develop into late adolescence. Ross et al. (2011) suggested that this developmental trajectory reflects not only increasing exposure to audiovisual speech in noisy environments but also the maturation of other perceptual and cognitive abilities that emerge concurrently with the development of audiovisual integration skills.

The ability of children to perceive speech has also been investigated when the speech signal is distorted, rather than masked. Noise vocoding attempts to simulate speech as heard via a cochlear implant in listeners with normal hearing, degrading its spectral (frequency) content while retaining its temporal envelope dynamics (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Arguably, this body of research has been integral to understanding speech perception difficulties experienced by cochlear implant users, as well as how these difficulties can be potentially remediated. Critically, across a number of studies, it has been shown that 5- to 7-year-old children with normal hearing require a greater spectral resolution (more frequency bands) in the signal to identify noise-vocoded speech, compared with older (10- to 12-year-old) children and adults (e.g., Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Newman & Chatterjee, 2013; Vongpaisal, Trehub, Schellenberg, & van Lieshout, 2012). Eisenberg et al. (2000) attributed this finding to young children's inability to use all of the sensory information available to them, as well as being due, in part, to the incomplete development of linguistic and cognitive abilities. Nevertheless, whether additional visual cues improve children's perception of degraded (e.g., noise-vocoded) speech is one key area that is yet to be explored. Based on evidence showing that audiovisual integration in noise develops as a function of age (Barutchu et al., 2010; Ross et al., 2011), the same may be true when the auditory speech signal is spectrally degraded.

The main objective of the current study was to explore the extent to which children with normal hearing benefit from visual information when it is matched to an auditory speech signal degraded by noise vocoding (as opposed to when the fidelity of the signal is reduced by noise masking)—a simulation of speech distortion as heard via a cochlear implant based on our understanding of perceptual mechanisms. Understanding the development of audiovisual enhancement afforded by access to visual information when the auditory signal is degraded in this way provides a promising line of research, potentially providing novel methods for assessing hearing and language development in pediatric users of cochlear implants, as well as assisting in the development of new treatment and intervention strategies. In listeners with normal hearing, we expected to replicate existing findings that young children (between 4 and 7 years of age) require greater spectral resolution (i.e., more frequency bands) to identify the speech signals at the same level as adults and older children ages 8–11 years (Eisenberg et al., 2000). In addition, if the same underlying mechanisms support audiovisual integration when the speech signal is either masked by background noise or is itself degraded by noise vocoding, we expected the following: On the basis of the finding that children benefit less from accompanying visual information when auditory speech is presented in noise (e.g., Barutchu et al., 2010; Ross et al., 2011), young children would show a reduced amount of benefit in comparison to both older children and adults when noise-vocoded sentences were presented simultaneously with corresponding mouth movements (Petersen & Posner, 2012; Posner & Petersen, 1990).

Method and Materials

Participants

Eighty-two native English-speaking children (38 boys, 44 girls) and 15 adults (2 men, 13 women) took part in the current study. Children ages 4–11 years were recruited through the University of Nottingham's seventh annual Summer Scientist event. At the event, advertised via local schools, newspapers, and radio stations, children visit the university to participate in a range of scientific studies in a fun and interactive manner (for more information, see http://www.summerscientist.org/). On the basis of parental report, children did not experience learning difficulties or problems with their hearing. Children were separated into four groups on the basis of age: 4–5 years (n = 15; 8 boys, 7 girls; M = 4.65), 6–7 years (n = 30; 12 boys, 18 girls; M = 6.81), 8–9 years (n = 24; 12 boys, 12 girls; M = 8.76), and 10–11 years (n = 13; 6 boys, 7 girls; M = 10.68).

Adult listeners, ages 18 to 26 years, were recruited separately via posters from the University of Nottingham student population and the general public and were paid an inconvenience allowance for their participation. All adult participants had normal hearing (pure-tone thresholds $\leq$ 20 dB HL across 0.125–8 kHz octaves; British Society of Audiology, 2011).

Written consent was obtained from the participant (adults) or a responsible caregiver (children), with each child giving verbal assent to participate. The study was conducted in accordance with and the approval of the University of Nottingham's School of Psychology Research Ethics Committee (children) and the Nottingham University Hospitals Research Ethics Committee (adults).

Stimuli

The stimuli used in the present investigation were adapted from the Children's Co-ordinate Response Measure (CCRM; Rosen, 2011). The CCRM presents listeners with a target sentence taking the form, "Show the dog where the [*color*] [*digit*] is." For each sentence, there is a total of 48 color–number combinations, which are selected at random on a trial-by-trial basis from a corpus consisting of six possible colors (blue, black, green, pink, red, white) and eight digits (1–9, excluding 7). The target response required of a listener is to identify the color and digit from the sentence.

For the current study, each sentence was filmed three times in black and white when spoken by a female in a monotone voice (at an F0 of approximately 212 Hz), using a Panasonic AG-HMC41E digital camcorder, in a sound-attenuated booth with a plain background. Video files were subsequently converted into an audio video interleaved (avi) format using Adobe Premiere Pro CS6 (Adobe Systems Software) and cropped so that only the mouth and lower jaw were visible in Final Cut Express 4.0 (Apple Mac, PowerMac, Mac OS X). The sound portion of each video file was also converted into uncompressed waveform audio file (wav) format to serve as the auditory stimulus. Noise vocoding was performed on each audio file using the methods akin to the published standards of Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995): Each sentence was noise vocoded with 1–25 frequency bands using TigerCIS software (Tigerspeech Technology, Qian-Jie Fu, House Ear Institute). The analysis input frequency range was 200–7000 Hz with a roll-off of 24 dB per octave. The signal was split into frequency bands using bandpass filtering (Butterworth filters, 24 dB per octave roll-off), and the envelope of each band was extracted using half-wave rectification and low-pass filtering (400 Hz cutoff). The envelope derived from each band was used to amplitude-modulate a bandpass carrier with the same band width as the original signal band. The resulting modulated noises were combined at equal amplitude ratios to create the final noise-vocoded stimuli.

General Procedure

Testing was completed in a sound-attenuated booth. The task was administered via custom software written in MATLAB and Statistics Toolbox Release R2010a (MathWorks) at a sound level of 65 dB SPL. For the auditory-only (AO) condition, the speech sounds were accompanied by a still frame of the speaker with closed lips and a neutral expression. For the audiovisual (AV) condition, the speech sounds were presented with a video recording of the talker so that both auditory and visual information were shown simultaneously. Visual material was displayed in the center of a Stimulus screen, while auditory stimuli were presented diotically via a Fast Track Pro USB audio interface (M-Audio, inMusic Brands) and Sennheiser HD 25-I headphones.

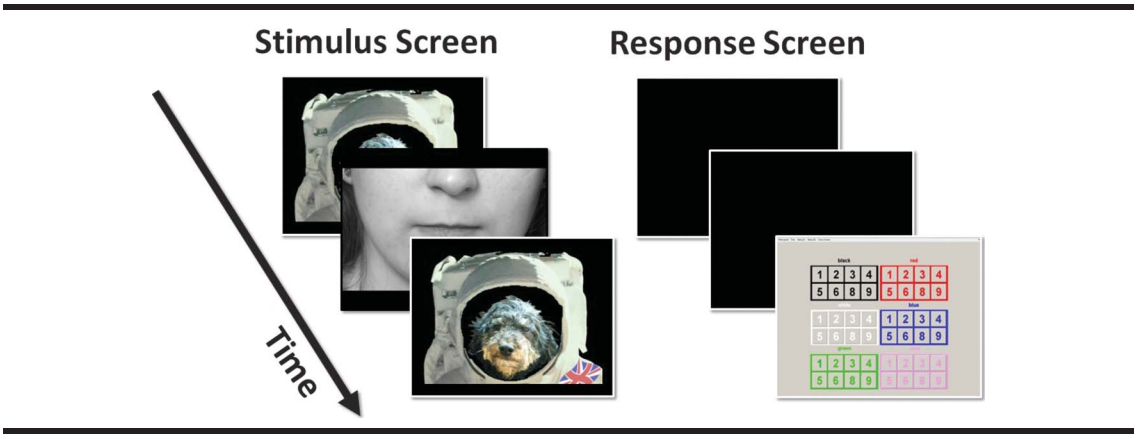
At the start of the experiment, participants were seated in the booth and told that they would be playing a game that involved saving the experimenter's dog, Lenny, who was lost in space. To help Lenny, they were told that they would have to interpret "alien" (i.e., spectrally degraded or noise-vocoded) speech, which would always take the form of a sentence such as "Show the dog where the green three is." Participants were instructed to accurately identify the color and number presented within each sentence in order to help Lenny land safely on earth. Before commencing, the experimenter confirmed with each child that he or she could correctly identify all colors and numbers that would be presented visually on the response screen.

At the start of each trial, participants were asked to look directly at the Stimulus screen displaying a picture of Lenny whenever the warning sound (Computer Noise 6, http://www.mediacollege.com/downloads/sound-effects/star-trek/tos/) was heard (see Figure 1). This was then followed by presentation of the target sentence, which was approximately 5 s in duration. The target sentence was immediately followed by the same picture of Lenny presented at the start of the trial. Participants were then required to make a response via the touch screen placed directly adjacent to the Stimulus screen, displaying all possible response options. There was no time limit in which to respond, and following a response the next trial commenced automatically.

AV and AO conditions were presented in separate blocks (two each), the order of which was counterbalanced for all possible orders across participants. A demo of five trials was administered before the first test block for each presentation condition to familiarize participants with the task requirements. After a response had been made, feedback was given by highlighting the correct response in purple on the response screen, as well as presenting the clear unprocessed version of the stimulus, followed by a repeat presentation of the original vocoded sentence. All participants were required to identify all demo trials correctly before progressing to the test phase.

During each trial of the test phase, the number of bands was adaptively varied, starting with 25 bands, and was reduced in steps of four bands according to a one-down, one-up staircase procedure. Following the first incorrect response, the number of frequency bands was adaptively varied using a three-down, one-up staircase procedure, targeting 79.4% correct on the psychometric function (Levitt, 1971). The number of bands was reduced by one band following three correct responses and increased by one band following an incorrect response. Feedback was not provided, and the adaptive track was terminated after 25 trials had elapsed. At least one threshold estimate was completed

Figure 1. A schematic representation of the sequence of the stimuli presented for each trial. The visual stimulus was always presented via a Stimulus screen, with the adjacent touch screen placed to the right. The response options were visible only immediately after the target sentence had been presented and disappeared once an option had been selected.



for each presentation mode. If time allowed, two estimates were completed, with the threshold determined as the average of both thresholds obtained.

The experimenter remained in the booth throughout the experiment, prompting participants to focus on the relevant screen whenever this was necessary. The experimental procedure lasted approximately 20 min.

Data Analysis

Rather than simply measuring the percent correct under a constant number of frequency bands, we parametrically manipulated the amount of degradation, and hence speech intelligibility, by varying the number of bands used for vocoding. Adaptive techniques are commonly used in auditory psychophysics, including speech perception (Brand & Kollmeier, 2002; Kollmeier & Wesselkamp, 1997; Ozimek, Warzybok, & Kutzner, 2010; Wagener, Brand, & Kollmeier, 1999a, 1999b; Wagener, Kühnel, & Kollmeier, 1999). They avoid ceiling and floor effects inherent in measuring percent correct performance at a fixed signal-to-noise ratio and equate the subjective level of difficulty for all listeners. Measuring performance in this way allowed us to determine a precise perceptual threshold (in terms of number of frequency bands) for each individual that could be meaningfully compared across all age groups.

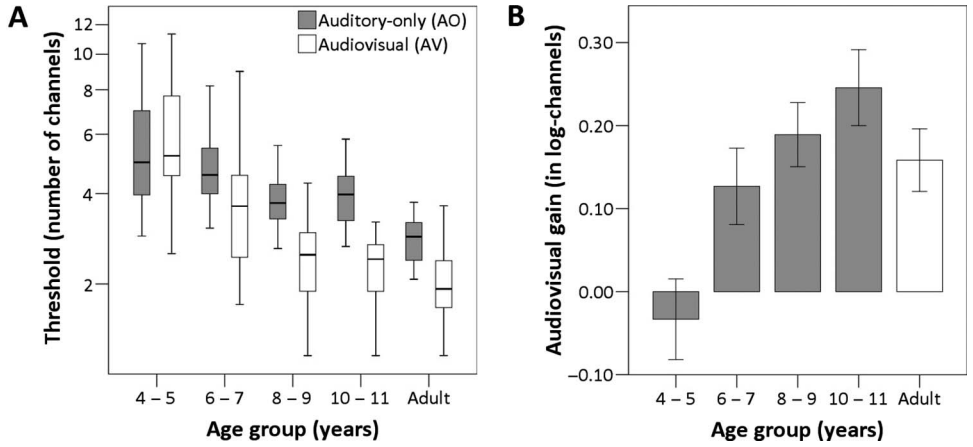
The data were log-transformed because the assumption of equal variances was violated (Levine's test: $p \leq .001$). Following this transformation, data were both normal ($p \geq .1$) and of equal variance ($p \geq .35$). Logistic psychometric functions were fitted to the log-transformed number of frequency bands from each adaptive track using the Psignifit toolbox (Wichmann & Hill, 2001). Thresholds were estimated as the 79.4% correct point on this function. Tracks where the optimization procedure did not adequately fit the data (i.e., when the fitted slope was negative or when the fitted value was outside the measured range) were discarded (see Amitay, Irwin, Hawkey, Cowan, & Moore, 2006)—occurring for 8.9% of all threshold estimates. Twelve children were subsequently excluded from the analysis on the basis that no threshold estimate was available for the AO ($n = 5$) and/or AV ($n = 8$) presentation conditions. One outlying participant within the 6- to 7-year-old age group was also excluded because of highly inconsistent threshold estimates. Excluding this participant did not alter the main result but reduced variability considerably. A total of 69 children were consequently included in the overall analysis (4–5 years, $n = 11$; 6–7 years, $n = 23$; 8–9 years, $n = 22$; 10–11 years, $n = 13$).

To establish whether performance differed between presentation conditions according to age, we compared thresholds for the AO presentation condition with those obtained in the AV condition by using a 2 × 5 mixed analysis of variance, with presentation mode (AO, AV) as the within-subjects factor and age group (4–5, 6–7, 8–9, 10–11, adults) as the between-subjects factor.

Results

The log-transformed thresholds for 79.4% correct speech identification (see Figure 2A) decreased as a function of age, $F(1, 79) = 11.8$, $p < .001$, $\eta_p^2 = .37$. Thresholds for AV presentation were significantly lower than for AO presentation, $F(1, 79) = 43.6$, $p < .001$, $\eta_p^2 = .36$. Moreover, the interaction between age group and presentation condition was significant, $F(4, 79) = 3.99$, $p = .005$, $\eta_p^2 = .17$, suggesting that the benefit conferred by AV compared with AO changed with age. We explored this interaction by plotting the AV gain as the difference between (the log-transformed) AV and AO thresholds (Ross et al., 2011; see Figure 2B). Pairwise comparisons revealed a significant difference in AV gain between the 4- to 5-year-old group and all other age groups ($p \leq .02$), with no difference between the remaining age groups ($p \geq .2$). Furthermore, Holm–Bonferroni-corrected

Figure 2. Threshold estimates by age group. A: Box plot showing the number of frequency bands required to achieve accurate speech identification in the auditory-only (AO) and audiovisual (AV) presentation conditions. The dark line contained within each box is the median threshold. The bottom of each box indicates the 25th percentile, whereas the top represents the 75th percentile. T-bars denote the range of threshold values. B: The AV gain score, where the benefit of accompanying visual cues was measured by subtracting log-transformed AV threshold estimates from those obtained during AO presentation. Error bars denote ±1 standard error of the mean.



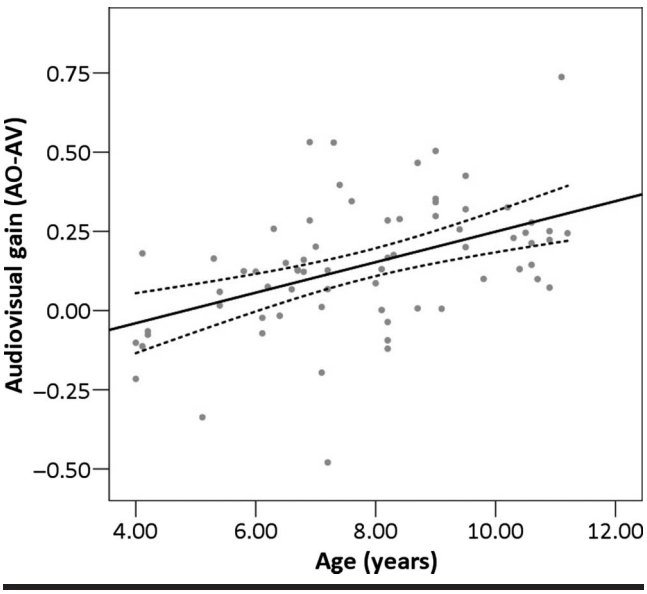
one-sample *t* tests on AV gain demonstrated that 4- to 5-year-old children showed no AV gain ($p = .51$), whereas all other children showed a significant AV gain ($p \leq .02$). Thus, whereas 4- to 5-year-old children did not benefit from accompanying visual information, 6- to 11-year-old children and adults performed better when the speaker's mouth movements were available.

Audiovisual gain in adults was (nonsignificantly) less than that observed in 8- to 9- and 10- to 11-year-old children (see Figure 2B). This is likely the result of a ceiling effect, whereby the adult listeners were able to correctly identify the closed-set speech stimuli with a single frequency channel in the AV condition (see Figure 2A), thereby limiting the possible gain provided by accompanying visual information. The linear relationship between the AV gain measure and age was further explored in the child sample only (see Figure 3): Age significantly and positively correlated with AV gain ($r = .47$), predicting 22% of the variance, $F(1, 67) = 19.02$, $p < .001$. AV gain therefore increased in children as a function of age.

Discussion

The objective of the current study was to investigate the development of audiovisual integration in children from 4 to 11 years of age when the auditory signal was spectrally degraded. Audiovisual integration was defined in terms of the gain achieved in noise-vocoded speech perception when visual information accompanied the auditory signal. The data revealed two distinct findings: First, we replicated previous work (Eisenberg et al., 2000; Newman & Chatterjee, 2013; Vongpaisal et al., 2012) showing that 4- to 5-year-old children required greater spectral resolution (i.e., a greater number of frequency bands) to identify speech targets in comparison with older children (6–11 years) and adults. Second, we found that AV gain increased as a function of age. That is, although there was no AV gain in 4- to 5-year-old children, from 6 years of age children began to benefit from accompanying visual speech cues. A similar finding has also been shown when auditory speech is masked by external noise (Barutchu et al., 2010; Ross et al., 2007, 2011; Wightman et al., 2006), suggesting that similar developmental processes underlying audiovisual integration might

Figure 3. Scatter plot showing the linear relationship between children's AV gain and age. The plot shows line of best fit (solid line) and 95% confidence intervals (dotted lines).



be involved when the auditory speech signal is spectrally degraded.

There is a long established literature showing that children are less susceptible than adults to the influence of visual information when it does not match the auditory speech signal (Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008). One possible explanation for the developmental shift observed around 6 years of age is that younger children do not attend to the visual input or are unable to process information from multiple channels presented simultaneously because of limitations on processing capacity (Huang-Pollock, Carr, & Nigg, 2002; Lavie, 2005, 2010). Because of task demands or possibly because of bias, young children may be more sensitive to what they hear (Welch & Warren, 1980). In the present study, constraints were imposed to ensure that children attended to the visual stimulus. The video clips were also cropped, making only the lips visible so that the child could not be distracted by other facial features. An alternative possibility is that although young children do attend to the visual information, they are poorer speechreaders compared with older children and adults, resulting in a smaller benefit when the speaker's articulations accompanied degraded speech (Hockley & Polka, 1994; Massaro et al., 1986; Sekiyama & Burnham, 2008).

The onset of mainstream schooling is one significant factor that might increase the influence of visual speech cues in children (Massaro et al., 1986). Although younger children may be exposed to challenging auditory environments (e.g., in preschool), understanding and acting on the auditory information becomes important when children start school (Shield & Dockrell, 2003, 2008); the successful processing of the teacher's instructions to direct behavior becomes crucial for academic success (Ames, 1992; Covington, 2000; Meece, Anderman, & Anderman, 2006). Consequently, learning to effectively attend to and integrate visual speech cues during the early school years may emerge at a time in development when the ability to plan goal-directed behavior is also developing. In support, the developmental trajectory we observed for audiovisual integration abilities appears to be similar to that found for executive functions, including planning, cognitive flexibility, goal setting, and information processing (Anderson, 2002; Konrad et al., 2005; Welsh, Pennington, & Groisser, 1991). In view of this, our findings suggest that school-age children should be encouraged to view a teacher's articulations in order to improve comprehension.

The ability to use visual speech cues to enhance speech perception has also been shown to improve educational achievement in children who are deaf or hearing impaired (Geers, 2002; Kyle & Harris, 2006); this further underscores the importance of encouraging the integration of both auditory and visual channels in this population. The current findings may be particularly relevant to our understanding of how speech perception abilities can be improved in pediatric cochlear implant users. The findings suggest that chronological age might be critical when assessing the efficacy of intervention and treatment programs that aim to facilitate audiovisual integration abilities in pediatric cochlear implant users. However, future studies of audiovisual integration in cochlear implant users must also consider, in addition to chronological age, the complex interaction between other factors that have already been shown to be associated with this ability, such as age of implantation and communication mode (e.g., total or oral; Bergeson, Pisoni, & Davis, 2003; Lachs et al., 2001).

The implications our findings may have for cochlear implant technology are limited by the restricted nature of this experiment. We showed that even the youngest children need as few as 10 channels to accurately perceive the speech in the closed set we used—far fewer than the 22 channels often afforded by modern cochlear implant devices. Older children require even fewer, with supplementary visual cues further reducing the requisite number of channels. However, the speech stimuli used in the current study do not reflect a natural listening context. First, we used a small, closed stimulus set. This limitation is apparent in the ceiling effect seen in some older children as well as in adults: The AV task could often be done even with a single channel. The number of channels required has been shown to increase in both children and adults with normal hearing when a larger stimulus set was presented (Eisenberg et al., 2000). Second, the greatest challenge to speech perception with a cochlear implant is the presence of background noise, which was entirely absent in this experiment. Even adult listeners with normal hearing require more spectral bands to perceive vocoded speech in noisy listening conditions (Friesen, Shannon, Baskent, & Wang, 2001). To generalize these findings, future research should use larger stimulus sets, preferably open, and presented in vocoded background noise.

Taking these findings together, the current study demonstrates that 4- to 5-year-old children with normal hearing may not benefit from additional visual cues to enhance degraded speech perception. Furthermore, the gain from seeing a speaker's lip movements during the perception of spectrally degraded speech appears to develop progressively in children, from 6 to 11 years of age. As such, it is likely that this developmental trajectory not only reflects increasing exposure to degraded speech but that audiovisual integration skills develop simultaneously with higher-order, cognitive abilities. These findings have potential repercussions for facilitating academic success in children with normal hearing by encouraging audiovisual integration. Furthermore, this research may assist understanding of the how hearing and language skills can be assessed and potentially remediated in child users of cochlear implants.

Acknowledgments

This research was supported by the Medical Research Council, United Kingdom (Grant U135097130). We thank Karen Banai, Lorna Halliday, and Antje Heinrich for their comments on a draft.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261–271.

Amitay, S., Irwin, A., Hawkey, D. J., Cowan, J. A., & Moore, D. R. (2006). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners. *The Journal of the Acoustical Society of America, 119,* 1616–1625.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8,* 71–82.

Barutchu, A., Crewther, S. G., Fifer, J., Shivdasani, M. N., Innes-Brown, H., Toohey, S., … Paolini, A. G. (2011). The relationship between multisensory integration and IQ in children. *Developmental Psychology, 47,* 877–885.

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

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *Volta Review, 103,* 347–370.

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

Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience, 21,* 1790–1804.

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111,* 2801–2810.

British Society of Audiology. (2011). *Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking*. Available at http://www.thebsa.org.uk

Burnham, D., & Dodd, B. (1996). Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 103–114). New York, NY: Springer.

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

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51,* 171–200.

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

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

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110,* 1150–1163.

Geers, A. E. (2002). Factors affecting the development of speech, language, and literacy in children with early cochlear implantation. *Language, Speech, and Hearing Services in Schools, 33,* 172–183.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103,* 2677–2690.

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America, 96,* 3309.

Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults. *Developmental Psychology, 38,* 363–375.

Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *The Journal of the Acoustical Society of America, 102,* 2412–2421.

Konrad, K., Neufang, S., Thiel, C. M., Specht, K., Hanisch, C., Fan, J., … Fink, G. R. (2005). Development of attentional networks: An fMRI study with children and adults. *NeuroImage, 28,* 429–439.

Kyle, F. E., & Harris, M. (2006). Concurrent correlates and predictors of reading and spelling achievement in deaf and hearing school children. *Journal of Deaf Studies and Deaf Education, 11,* 273–288.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing, 22,* 236–251.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences, 9,* 75–82.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19,* 143–148.

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

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 1431–1436.

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

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

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia, 50,* 762–776.

McGurk, H., & MacDonald, J. (1976, December 23–30). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57,* 487–503.

Newman, R., & Chatterjee, M. (2013). Toddlers' recognition of noise-vocoded speech. *The Journal of the Acoustical Society of America, 133,* 483–494.

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49,* 444–454.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35,* 73–89.

Posner, M. L., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Rosen, S. (2011, June). *The complexities of understanding speech in background noise*. Paper presented at the First International Conference on Cognitive Hearing Science for Communication, Linköping, Sweden.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59,* 347–357.

Ross, L. A., Molholm, S., Blanco, D., Gomez Ramirez, M., Saint Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience, 33,* 2329–2337.

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

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research, 1188,* 87–99.

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

Sánchez-García, C., Alsius, A., Enns, J. T., & Soto-Faraco, S. (2011). Cross-modal prediction in speech perception. *PloS One, 6*(10), e25198.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 18748–18750.

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

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995, October 13). Speech recognition with primarily temporal cues. *Science, 270,* 303–304.

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics, 10,* 97–116.

Shield, B. M., & Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *The Journal of the Acoustical Society of America, 123,* 133–144.

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

Summerfield, Q., MacLeod, A., McGrath, M., & Brooke, M. (1989). Lips, teeth, and the benefits of lipreading. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 223–233). Amsterdam, the Netherlands: Elsevier Science.

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., & van Lieshout, P. (2012). Age-related changes in talker recognition with reduced spectral cues. *The Journal of the Acoustical Society of America, 131,* 501–508.

Wagener, K., Brand, T., & Kollmeier, B. (1999a). Development and evaluation of a German sentence test: Part 2. Optimization of the Oldenburg Sentence Test. *Zeitschrift Fur Audiologie, 38,* 44–56.

Wagener, K., Brand, T., & Kollmeier, B. (1999b). Development and evaluation of a German sentence test: Part 3. Evaluation of the Oldenburg Sentence Test. *Zeitschrift Fur Audiologie, 38,* 86–95.

Wagener, K., Kühnel, V., & Kollmeier, B. (1999). Development and evaluation of a German sentence test: 1. Design of the Oldenburg Sentence Test. *Zeitschrift Fur Audiologie, 38,* 4–15.

Walden, B. E., Prosek, R. A., & Worthington, D. W. (1974). Predicting audiovisual consonant recognition performance of hearing-impaired adults. *Journal of Speech and Hearing Research, 17,* 270–278.

Walden, B. E., Prosek, R. A., & Worthington, D. W. (1975). Auditory and audiovisual feature transmission in hearing-impaired adults. *Journal of Speech and Hearing Research, 18,* 272–280.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638–667.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7,* 131–149.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: 1. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

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